Uncertainties in Land Change Modeling

DISSERTATION

zur Erlangung des akademischen Grades: Dr. rer. nat.

im Fach Geographie

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
Carsten Krüger

Präsident der Humboldt-Universität zu Berlin

Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät

Prof. Dr. Elmar Kulke

Gutachter/Gutachterin:

1. Prof. Dr. Tobia Lakes
2. Prof. Dr. Daniel Müller
3. Prof. Dr. Rüdiger Schaldach

Tag der Verteidigung: 28.04.2016

Danksagung

Viele Wochen, Monate, sogar Jahre sind verstrichen, seitdem sich die ersten Ideen entfalteten. Sie formten sich zu Bits und Bytes und ergeben nun diese zu Papier gebrachte Arbeit. Aber nicht nur eine Dissertation formt sich in einer so langen Zeit, auch der Doktorand selbst wird durch sie nachhaltig geprägt. Er sammelt sein Rüstzeug zusammen und betritt einen unbekannten Pfad um sich auf eine lange Reise zu begeben. Viele Leute waren daran beteiligt, die richtigen Abzweigungen auf diesem Pfad zu finden. So wurden viele Unwägbarkeiten umschifft oder manchmal auch gezielt auf sich genommen.

Daher möchte ich mich von ganzem Herzen bei allen bedanken, die mich auf diesem Weg unterstützten und insbesondere bei:

Tobia Lakes, für den Kompass, den sie mir mit auf den Weg gab und mir zeigte wie man ihn bedient.

Dennis Funke, der mich auf meinem Weg begleitete und mir oft meinen schweren Rucksack abnahm.

Dem gesamten Team von CarBioCial, für einen enormen Fundus an Landkarten, mit denen ich in neue Regionen vorstoßen konnte.

Christian Schuster und Florian Gollnow für ihre Grubenlampe, die oftmals Licht in das Dunkel brachte.

Daniel Müller, Rüdiger Schaldach und Ulf Leser, für ihr Moskitonetz, um mich vor unliebsamen Angreifern zu schützen.

Robert Gilmore Pontius, der mir das Entdecken und Lesen von Spuren beibrachte.

Dem Metrik-Team und Joachim Fischer, für ihr Fernglas, mit dem man neue Perspektiven einnehmen konnte.

Der DFG, die mir die nötige Zeit verschaffte, um auf diese lange Reise zu gehen.

Dagmar Wörister und Gabriele Graichen, für ihre Armbanduhr, die mir zeigte, ob ich den richtigen Ort zur richtigen Zeit erreicht hatte.

Jim Curtiss, Mary Dellenbaugh und Stephan Bahn, die dabei halfen meine Reise für den Lesenden festzuhalten.

Andreas Rabe, Benjamin Jakimow, Cheong Yoon Ling, Christian Geiß, Doris Dransch, Johannes Schreyer, Philippe Ruffin und dem gesamten Land System Science Cluster, die mit mir Alternativrouten diskutierten.

Meiner Familie, die mich motivierte, den Weg weiterzugehen.

Meiner Frau Jördis, für alles.

Abstract

Human influence has led to substantial changes to the Earth's surface. One example of this is the deforestation of tropical rainforest to satisfy the increasing demand of the global economic markets for soy and beef. Land change models are widely applied to analyze such processes and to give recommendations for decision-making. These models are used to investigate which factors are relevant for current land change and which future developments are likely. Land change models are affected by uncertainties which have to be taken into account when interpreting their results. However, approaches which examine different sources of uncertainty with regard to their interdependencies and their influence on projected land change are rarely applied. The first objective of this thesis is therefore to develop a systematic approach which identifies major sources of uncertainty and the propagation to the resulting land change map. Another challenge in land change modeling is the estimation of the reliability of land change predictions when no reference data are available. This issue is frequently addressed with a qualitative comparison of different land change scenarios. In this approach, the level of uncertainty remains unknown. The second objective of this thesis is therefore to quantify the uncertainty about future land change. Bayesian Belief Networks were identified as a useful technique to reach the first objective. Moreover, the modeling steps of "model structure definition", "data selection" and "data preprocessing" were detected as relevant sources of uncertainty. This thesis additionally observed that the uncertainty in the modeling process and the accuracy of the model output are not substantially interrelated. To address the second research objective, a set of measures based on probabilities were developed. They quantify uncertainty by means of a single predicted land change map without using a reference map. It is additionally possible to differentiate uncertainty into its spatial and quantitative components by means of these measures. This is especially useful in spatial applications such as land change modeling. However, even a certain model can be wrong and therefore useless. Therefore, an approach is suggested which estimates the relationship between disagreement and uncertainty in known time steps. This relationship can be used to assess the reliability of land change predictions when only the quantification of uncertainty is possible. Apart from the quantification of uncertainty based on one map, another approach which is based on the comparison of different land change predictions was developed in this dissertation. The approaches give important information for understanding the reliability of possible future development paths. Moreover, they are transferrable to other spatial research disciplines which are based on probabilities.

Zusammenfassung

Der Einfluss des Menschen verändert die Erdoberfläche in gravierendem Maße. Ein Beispiel ist die Entwaldung von tropischen Regenwäldern um den steigenden Bedarf der globalen Märkte nach Soja und Rindfleisch zu decken. Die Anwendung von Landnutzungsmodellen ist etabliert, um derartige Prozesse zu analysieren und um Handlungsempfehlungen für Entscheidungsträger zu geben. Mit diesen Modellen wird untersucht, welche Faktoren für auftretende Änderungen der Landschaft wesentlich sind und welche zukünftigen Entwicklungen wahrscheinlich sind. Landnutzungsmodelle stehen unter dem Einfluss von Unsicherheiten, welche beim Interpretieren der Ergebnisse berücksichtigt werden müssen. Dennoch gibt es wenige Ansätze, die unterschiedliche Unsicherheitsquellen mit ihren Interdependenzen untersuchen und ihre Auswirkungen auf die projizierte Änderung der Landschaft analysieren. Aus diesem Grund ist das erste Ziel dieser Arbeit einen systematischen Ansatz zu entwickeln, der wesentliche Unsicherheitsquellen analysiert und ihre Fortentwicklung zur resultierenden Änderungskarte untersucht. Eine andere Herausforderung in der Landnutzungsmodellierung ist es, die Eignung von Projektionen abzuschätzen wenn keine Referenzdaten vorliegen. Dieses Problem wird häufig adressiert indem verschiede Szenarien in qualitativer Weise miteinander verglichen werden. Dabei bleibt die Höhe der Unsicherheit unbekannt. Das Quantifizieren von Unsicherheiten in zukünftigen Änderungen der Landschaft ist aus diesem Grund das zweite Ziel dieser Arbeit. Bayes'sche Netze wurden als eine vielseitige Methode identifiziert, um das erste Ziel zu erreichen. Darüber hinaus wurden die Modellierungsschritte „Definition der Modellstruktur", „Auswahl der Eingangsdaten" und „Weiterverarbeitung der Daten" als wesentliche Unsicherheitsquellen identifiziert. In dieser Dissertation wurde zusätzlich beobachtet, dass die Unsicherheit im Modellierungsverlauf und die Genauigkeit des Modellergebnisses nicht zwangsläufig voneinander abhängig sind. Um das zweite Ziel zu adressieren wurde eine Auswahl an Maßzahlen entwickelt. Diese quantifizieren Unsicherheit mit Hilfe einer projizierten Änderungskarte und ohne den Vergleich mit Referenzdaten. Mit diesen Maßzahlen ist es zusätzlich möglich zwischen quantitativer und räumlicher Unsicherheit zu unterscheiden. Vor allem in räumlichen Anwendungen wie der Landnutzungsmodellierung ist diese Möglichkeit wertvoll. Dennoch kann auch ein absolut sicheres Modell gleichzeitig ein falsches und nutzloses Modell sein. Deswegen wird ein Ansatz empfohlen, der die Beziehung zwischen Unsicherheit und Genauigkeit in bekannten Zeitschritten schätzt. Diese Beziehung kann im Folgenden genutzt werden um die Eignung von Landschaftsprojektionen zu analysieren, wenn nur die Quantifizierung von Unsicherheiten möglich ist. Abgesehen von dieser Methodik wurde ein weiterer Ansatz in dieser Dissertation entwickelt. Dieser basiert auf einem Vergleich verschiedener Projektionen. Die entwickelten Ansätze geben wichtige Informationen um die Eignung von möglichen Entwicklungspfaden zu verstehen. Darüber hinaus sind sie übertragbar auf andere räumliche Forschungsströmungen, die auf Wahrscheinlichkeiten basieren.

Contents

List of Figures

List of Tables

List of Equations

List of Acronyms

AUC	Area under the curve
BBN	Bayesian Belief Network
CarBioCial	Carbon sequestration, biodiversity, and social structures in Southern Amazonia
DNIT	Departamento Nacional de Infraestrutura de Transportes
FAO	Food and Agriculture Organization of the United Nations
GDP	Gross Domestic Product
GRK	Deutsche Forschungsgemeinschaft - German Research Foundation
IBGE	Instituto Brasileiro de Geografia e Estatística
INPE	Insituto Nacional de Pesquisas Espaciais
IPCC	Intergovernmental Panel on Climate Change
IPEA	Instituto de Pesquisa Econômica Aplicada
Land change	Land use and land cover
METRIK	Modellbasierte Entwicklung von Technologien für selbstorganisierende dezentrale Informationssysteme im Katastrophenmanagement
MMA	Ministerio do Meio Ambiente
MT	Ministerio do Transportes
PRODES	Projeto de Monitoramento do Desmatamento na Amazônia Legal por Satélite
ROC	Receiver Operating Characteristics
SISCOM IBAMA	Sistema Compartilhado de Informações Ambientais do Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis
SRTM	Shuttle Radar Topography Mission

Mathematical symbols

$(1 - p_{ki})$	Probability for future "land persistence" in pixel i following projection k
a, b	Estimated regression coefficients
AD	Allocation disagreement
APD	Allocation probability disagreement
AU	Allocation uncertainty
CP	Sum of pixels with predicted "land change" in projection 1 and predicted "land persistence" in projections 2
cr	Correct rejection
D_i	Disagreement (quantity, allocation or total) in a municipality i
e_i	Residue
fa	False alarm
FA	Fraction of false alarms and the total number of modeled pixels
h	Hit
H(X)	Entropy of a Variable X
m	Miss
M	Fraction of misses and the total number of modeled pixels
Mean PCP	Average of $p_{1i} \cdot (1 - p_{2i})$ over all pixels
Mean PPC	Average of $(1 - p_{1i}) \cdot p_{2i}$ over all pixels
MI(Y;X)	Mutual information of the variable X and Y
N	Number of pixels
N_{tot}	Total number of nodes in a BBN
N_{un}	Number of such uncertain nodes in a BBN
$P(X_1,\ldots,X_n)$	A joint probability distribution
PC	Sum of pixels with predicted "land persistence" in projection 1 and predicted "land change" in projections 2
PCP	Probability of change in projection 1 vs. persistence in projection 2
PF	probability false alarms: sum of probabilities of all cells to be a false alarm, divided by the total number of modeled pixels
p_{ki}	Probability for future "land change" in pixel i following projection k
PM	Probability misses: sum of probabilities of all cells to be a miss, divided by the total number of modeled pixels
PPC	Probability of persistence in projection 1 vs. change in projection 2
QD	Quantity disagreement
QPD	Quantity probability disagreement
QU	Quantity uncertainty
R	Correlation coefficient
R^2	Coefficient of determination
SU	Structure uncertainty in a BBN
TD	Total disagreement
TPD	Total probability disagreement
TU	Total uncertainty
U_i	Uncertainty (quantity, allocation or total) in a municipality i
X, Y	Given variables
x, y	Possible values of X and Y

Chapter I:
Introduction

1 Analyzing land change with models

The Earth's surface has changed over time. The rate of change has recently been increasing, above all as land uses have become more and more adapted to human needs. The accelerating change process started in industrialized countries, where natural or semi-natural ecosystems have been converted to urban areas, croplands or meadows (Foley *et al.* 2005). Multiple land use interests are motivations for humans to change the land, e.g. to enhance or to optimize their food and shelter supply. Today's most rapid changes can be detected in emerging countries such as China or Brazil (Diniz *et al.* 2013).

The land change processes addressed in this thesis imply changes to both land cover and land use. Land cover is related to the biophysical characteristics of the land surface, whereas land use refers to the specific anthropogenic utilization of the land cover in a certain area (Lambin and Geist 2006; Kim 2010). A single land cover type can be used in different ways by humans. While land cover can be detected by means of satellite imagery, land use cannot be derived directly from this source. Land use and land cover changes can lead to severe biophysical consequences such as increasing atmospheric CO_2 level, biodiversity loss, increasing nutrient inputs in soils, or an increasing exposure to natural hazards. Moreover, land change processes have socioeconomic impacts. They can, for example, lead to increasing trade activities or a higher attraction for labor (Asselen and Verburg 2013).

Land changes can be seen as both a cause and a consequence of human activities and are a part of the human-environment system (Verburg *et al.* 2010). Humans are on the one hand dependent on the vitality of the global ecosystem, and influence the vitality of this system on the other. The challenge is to obtain the necessary benefits from the ecosystem while simultaneously mitigating the negative impacts of land change activities (Foley *et al.* 2005). Therefore, it is crucial to understand the interactions within the land system (Geist and Lambin 2002; Pijanowski *et al.* 2002).

Gaining insight into the land system is not a straightforward process. Human-induced land change processes are complex, due to feedbacks, loops, indirect effects or nonlinear behavior in the coupled human-environment system (Arima *et al.* 2011; Sloan and Pelletier 2012; Hagos *et al.* 2014; Richards *et al.* 2014). For example, land systems can persist in a stable situation for a long time period, followed by rapid unforeseeable changes until a new stable situation is reached. This movement from one equilibrium to another is termed a regime shift (Müller *et al.* 2014). Another example of complexity is the self-intensification of changes due to feedbacks. To understand these complex processes, land change models are used.

These models are mainly applied for two purposes: to explain already occurring land change processes and to project land change in unknown time steps or in unknown regions. The first purpose is connected to the identification of the relevant drivers of land change. In comparison, the second purpose is related to the projection of new land change. This can be done by assuming that the past conditions will be the same in the future ("business as usual") (Kim 2010), or by including changes to the input variables to analyze the effect of changing circumstances (Soares-Filho *et al.* 2006; Batisani and Yarnal 2009). This second process is frequently termed scenario-based land change modeling (Guitierrez-Velez and Pontius 2012). The results of a land change model are valuable for example for urban and spatial planning processes. Using land change models, decision-makers could anticipate possible land change consequences, evaluate effective mitigating strategies and choose appropriate regions to implement spatial planning strategies (Pontius *et al.* 2001).

A variety of approaches to modelling land change have been developed which are based on different theoretical assumptions, have different aims and deliver different results. Following Brown *et al.* (2013), these approaches can be classified into five types:

a) Machine learning approaches (e.g. Lakes *et al.* 2009) focus on projecting future patterns based on past trends. Frequently, such spatial projections rely on information given by spatial variables.

b) Sector-based economic models (e.g. Hausman 2012) try to predict the demand for specific land-change classes in a certain region or sector.

c) Spatially disaggregated economic models (e. g. Müller *et al.* 2012) are econometric approaches, and are frequently based on micro-economic theory which predicts spatially explicit land change.

d) Agent-based models (e.g. Magliocca and Ellis 2013) express processes in an algorithmic form. They are frequently used to explore non-linear behavior and external shocks.

e) Cellular approaches: This type of land change model is a widely applied multi-scale approach. This approach combines a spatially explicit regional or local view and a non-spatial view of land change processes to project spatially explicit land change (Veldkamp and Lambin 2001; Diniz *et al.* 2013). Widely used land change models which follow this approach are the Land Use Scanner (Hilferink and Rietveld 1999), CLUE-S (Verburg *et al.* 2002), CLUMONDO (Asselen and Verburg 2013) and the LandSHIFT model (Schaldach *et al.* 2011). This thesis uses this approach as well, as illustrated in Figure I-1.

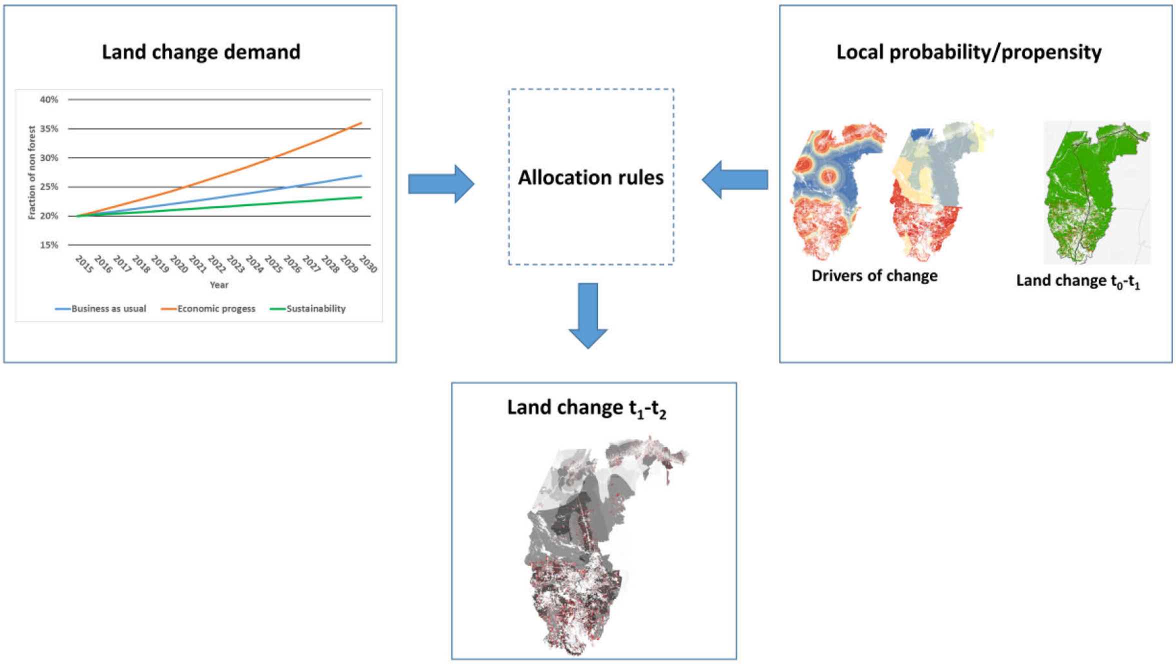


Figure I-1: Cellular land change modeling approach which is used in this dissertation

The motivation behind the cellular approach is that interactions within a land system are scale dependent. A certain driver could influence land change on a regional level but could be ineffective on a national level. For example, political incentives to encourage economic investments can be guided more efficiently on a smaller scale.

It is frequently assumed that the quantity of land change is driven by large scale economic drivers whereas regional and local characteristics determine the specific location of change (Verburg *et al.* 2004). The quantity is termed “land change demand” and is estimated for the whole study area in the first step. The demand can be derived as a stand-alone output from external models (Arsanjani *et al.* 2013), e.g. by time series analysis or economic models. Also, scenarios of land change are frequently used to quantify the demand. One example is the business as usual case which extrapolates past trends into the future. Alternatives are economic progress or sustainability scenarios. The IPCC (Intergovernmental Panel on Climate Change) storylines of the “Special Report on Emissions Scenarios” are also widely applied to construct different scenarios (Nakicenovic and Swart 2000). Reasonable assumptions are made as the scientific base for such scenarios. The second spatial scale represents the influence of local drivers. A continuous raster is produced by means of these drivers which quantifies whether a local entity, i.e. a pixel, is prone to future land change. Depending on the application and the author, the values of this raster are termed as probability, propensity, or suitability (Mas *et al.* 2013). The overlap between probabilities and propensities is that higher values characterize a higher chance of land change. In contrast to probabilities, propensities imply no information about the quantity of change. Suitability maps do not directly quantify the chance of one pixel to change. In this case, the change is additionally dependent on the suitability values of competitive classes. Following the development of the raster, an algorithm is defined which allocates the land change demand on the raster map according to local continuous values. The subsequent model output is a classified raster in which every pixel gets the class label of one land change transition.

2 Uncertainty in land change models

Uncertainty within the framework of land change models can substantially influence the reliability of these models. For example, uncertainty can disguise the existing risks of land change. Decision-makers tend to underestimate possible threats in this case. Moreover, uncertainty can affect the estimation of the effectiveness of mitigation strategies. For these reasons, land change models are evaluated. Among other evaluation subjects, land change modelers try to understand model uncertainties and validate projected land change patterns (Brown *et al.* 2013). The model validation covers two terms which are often used in the land change modeling literature: error and uncertainty. Messina *et al.* (2008) defined error as the difference between model estimates and measured reference data. Following these authors, uncertainty is a "degree of variability" which is attached to data values. Referring to these definitions, the terms are differentiated as follows in this thesis: Both error and uncertainty describe the deviation of the modeling output from the reality. The difference between the terms is the knowledge about reality. Error is the deviation from the known true value, whereas uncertainty is the deviation from an unknown true value. To quantify the error, a reference is necessary. In comparison, uncertainty tries to quantify the deviation without using any reference data. Another frequently used term is disagreement (Pontius and Millones 2011). In this dissertation, disagreement has the same meaning as error when modeled data and reference data are compared. Additionally, the term can be used when two model outputs are opposed. The term disagreement includes the meaning of error in the understanding of this study and is therefore used in the following chapters.

Uncertainties can have different sources. If land change modeling is understood as a process with different modeling steps, then a certain degree of uncertainty is connected to every modeling step (Walker *et al.* 2003). Three essential steps in land change modeling are: the definition of the model structure, the selection of appropriate data, and the preprocessing of these data. The first step deals with the identification of the relevant driving forces of land change. Moreover, relevant dependencies between the driving forces are included into the model. The second step is connected with the question of how many of the relevant drivers are used to construct an optimal land change model. Is a complex model with all of the relevant data the best choice or is a simple model with few divers the superior alternative? The best choice is a trade-off between using all the available information and a useful integration of the various sources. The third modeling step is concerned with the appropriate preparation of the required input data. One example is the question whether a simple Euclidean distance to a point of interest is the best possibility or whether a more sophisticated cost distance is more appropriate. Another example of a critical modelling issue is whether the original resolution of the data should be used or if it is beneficial to disaggregate the data into a finer resolution. Preprocessing work tries to approximate the input data to the imperfectly measurable reality; however, it is based on additional assumption and can include further uncertainty.

The amount of uncertainty in one of the mentioned steps is dependent on the uncertainty of the other steps (Krayer von Krauss *et al.* 2006). When trying to reduce uncertainty in the model, it is therefore helpful to cover these dependencies. Otherwise, the reduction of one source could lead to an increase of uncertainty in a second source. The uncertainty which arises in the modeling process propagates and aggregates the total model outcome uncertainty (Goldewijk and Verburg 2013). It is possible that uncertainty increases or diminishes through the modeling process. Although the importance and complexity of the interaction between different uncertainty sources is known in the land change community, approaches which systematically take this into account are rare (Goldewijk and Verburg 2013; Tayyebi, A.H. *et al.* 2014). Moreover, understanding about how the different uncertainty sources influence the reliability of the final model outcome is missing. The interaction of different uncertainty sources and their effect on the outcome accuracy is the first research gap which is addressed in this study.

Once a land change model is calibrated and applied, it delivers a land change output. This output can be an estimation for unknown areas (projection in space) or unknown time steps (projection in time). In both cases, the estimated land change is frequently given in a pixel based land change map. Therefore, a validation of the model output with reference data is completed in the majority of studies by summarizing agreeing and disagreeing pixels in a confusion matrix (Comber *et al.* 2012). A variety of accuracy measures, such as the overall disagreement, producer's accuracy, user' accuracy, Kappa, Fuzzy Kappa, or Receiver Operating Characteristics, are based on this matrix (e.g. Cohen 1960; Hagen-Zanker and Martens 2008; Diniz *et al.* 2013; Mas *et al.* 2013; van Vliet *et al.* 2013; Olofsson *et al.* 2014; Pontius and Parmentier 2014). Additionally, the quantity and allocation disagreement are widely used measures (Pontius and Millones 2011). They separate the total error into two parts. The first part refers to the disagreement of the quantity of the different land change categories in the model output and the reference data, while the second quantifies the spatial mismatch. This separation is especially useful when the modeling framework is based on the concept described above: e.g. specifying a regional land change demand on the one hand and estimating the local propensity of land change on the other hand.

Although a variety of accuracy and uncertainty measures have been developed, in land change modeling an underuse of these methods can be identified. The major reason given by different authors is data restrictions (Wassenaar *et al.* 2007; Jantz *et al.* 2014). A comparison with true reference data is not possible in a large number of land change applications, e.g. in future projections. The decision if a model is right or wrong in predicting land change at a certain area cannot be made in these cases. Alternatively, it can be useful to know if the model is certain or uncertain in its prediction. Therefore, several approaches have been developed and become common in the land change community, e.g. giving confidence intervals (Olofsson *et al.* 2013) or a probability distribution (Sangermano *et al.* 2012). The probability distribution can be summarized into uncertainty measures (Bastin *et al. 2012)* such as the mutual information criterion (Shannon and Weaver 1949). This probabilistic uncertainty concept is useful in pixel based land change modeling as well. However, a separation similar to the separation of disagreement into quantitative and spatial components is missing. This separation is particularly valuable in spatially explicit land change projections (Veldkamp and Lambin 2001) and is therefore identified as a second important research gap which is addressed in this dissertation.

3 Study area

The case study is situated in the Brazilian Amazon basin, in Pará and Mato Grosso state. The natural vegetation is tropical moist broadleaf forest and is home to a wealth of biodiversity (Eva *et al.* 2002). The climate is mainly the tropical wet climate (tropical rainforest, tropical monsoon, tropical savanna) following the Köppen-Geiger classification (Peel *et al.* 2007). A variety of rivers run through the region. The soils are predominantly acrisols and ferralsols with low fertility (Food and Agriculture Organization of the United Nations (FAO) 2015). The Brazilian Amazon is a sparsely populated area. The largest city in the chosen case study region is Itaituba with approximately 98.000 inhabitants (Instituto Brasileiro de Geografia e Estatística (IBGE) 2013). The BR-163 highway, which crosses the region from the south to north, is one of the most important roads in the area.

The region has been a hotspot of tropical deforestation in the last decades. Tropical deforestation is referred to as one of the most severe global land change processes (de Espindola *et al.* 2012). 168,873 km^2 of the Amazon rainforest were deforested between 2002 and 2014 (Insituto Nacional de Pesquisas Espaciais (INPE) 2015), an area equivalent to nearly half of the area of Germany. 69 % of this area is located in the case study states Pará and Mato Grosso. The land system and deforestation process has been investigated in several studies and a lot of knowledge is already available about the rates, locations, drivers and spatial determinants of deforestation. Substantial infrastructure development and colonization projects in the Amazon region initiated the extensive deforestation processes especially for the use of cattle ranching in the 1970s (Barona *et al.* 2010; Celentano *et al.* 2012). Subsequently, the increasing global demand for soybeans, beef, biofuels and timber have been identified as the most important drivers of the deforestation process (Arima *et al.* 2011; Richards *et al.* 2014). A high percentage of the forest clearings were initiated by middle and large-sized farmers to increase the production of agricultural goods (Fearnside 2008). This led to an expansion of areas dominated by agricultural monoculture and by cattle industry. Recently, the rate of deforestation has decreased due to the effect of political incentives such as the Soy Moratorium, a reduction of poverty, and increasing prices for soy and beef (Boucher *et al.* 2013; de Fries *et al.* 2013; Gollnow and Lakes 2014). Most of the development of the deforestation activities in Brazil have been driven by the influence of global agricultural markets in the last decades (Meyfroidt *et al.* 2013).

The region covers large parts of the Brazilian states Pára and Mato Grosso and includes 51 municipalities. The land change is derived from PRODES ("Projeto de Monitoramento do Desmatamento na Amazônia Legal por Satélite") data, the remote sensing based product of annual deforestation produced by the governmental institution INPE (Insituto Nacional de Pesquisas Espaciais (INPE) 2015). These data were widely applied in deforestation case studies of Brazil and can be seen as the standard product (Barona *et al.* 2010; Aguiar *et al.* 2012). Furthermore, a collection of 23 demographic, economic, biophysical and accessibility variables is used. These potential drivers of change were collected after a comprehensive literature review and in a close cooperation with the project CarBioCial ("Carbon sequestration, biodiversity, and social structures in Southern Amazonia", http://www.carbiocial.de).

4 Objectives and aims of this dissertation

Figure I-2 gives an overview of the framework of this dissertation. The framework comprises a sequence of steps relevant for most land use models which are related to the model calibration, validation and prediction. First, the land change model is calibrated with historic data. The calibration process is exposed to different sources of uncertainty. Second, the calibrated model is used to model present land change. The result is again influenced by uncertainty. A useful land change model can subsequently be used to project future land change. Uncertainty is a crucial part of the whole framework and propagates through the different steps. The first research objective addresses the links within the calibration as well as the transfer from the calibration to the validation. The second objective is related to the transfer from the validation to the prediction. The result is a comprehensive investigation of the influence of uncertainty in land change modeling. The sequence of steps which is connected within this investigation is presented below.

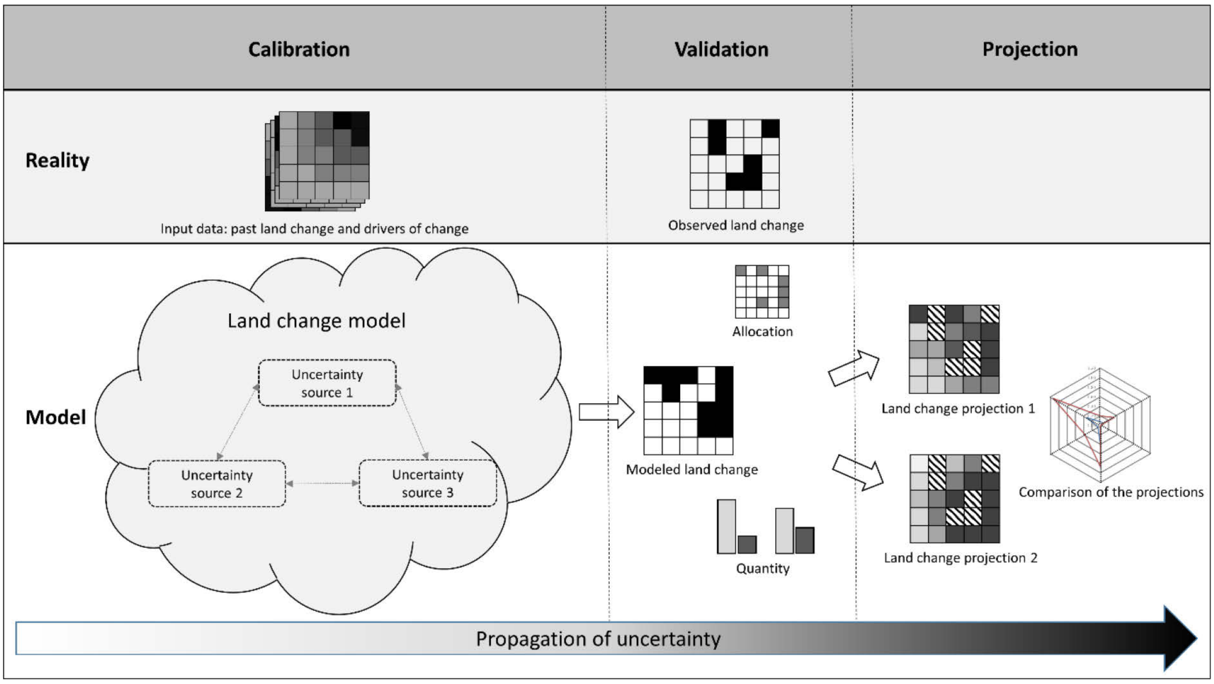


Figure I-2: Modeling and uncertainty framework of this dissertation

More specifically, the first research objective is: To develop an approach to systematically identify and analyze uncertainties in land change modeling

This objective is divided into the following specific steps:

- *Present an approach to quantify uncertainty in a land change model:* A probabilistic land change modeling approach is chosen. The probabilities imply uncertainty about reality and are used for different uncertainty measures. Therefore, the probability distribution is summarized into meaningful values.

- *Analyze the importance of different uncertainty sources which are connected to single modeling steps:* A set of imminent uncertainty sources is identified. Additionally, their isolated influence and the interactions between the sources are quantified.

- *Investigate how the uncertainty propagates throughout the modeling process and affects the accuracy of the model's outcome:* This thesis examines if important uncertainty sources of the calibration phase substantially influence the accuracy of the modeling results as well. Therefore, real change data are compared with modeled land change maps when the different modeling steps are modified.

The second research objective is: To develop methods to quantify uncertainty in land change projections which differentiate between spatial and quantitative uncertainty

This objective is addressed by the following four specific steps:

- *Separate the uncertainty of the modeling results into different components:* Land change modeling frequently combines different scales. Several models use the specified quantity of estimated change for a specific region or country and allocate this quantity on a finer scale. For that reason, it is useful to differentiate between quantity uncertainty and allocation uncertainty which are new measures developed in this thesis.

- *Quantify the relationship between uncertainty and disagreement in known time steps to allow conclusions about the future reliability of land change models:* It is only possible to assess disagreement if some kind of reference data is available. The two uncertainty measures developed in this thesis use no reference data; however, they cannot quantify the goodness-of-fit of a land change model. Even an absolutely certain model can be an unreliable and wrong model. Therefore, one approach is suggested to investigate if a relationship between uncertainty and disagreement exists in known time steps. Such a relationship could be a hint for the reliability of land change predictions of the same model when only the quantification of uncertainty is possible.

- *Compare different projections and quantify their difference to assess the uncertainty of future land change:* Land change projections are one possible future development. Only investigating single projections could give misleading indications about their probability to occur. This aim is therefore an alternative to the procedure presented in the previous research step. Different projections are taken into account to define an uncertainty range. If two projections based on reasonable assumptions are similar, then the uncertainty about the future development is low. Substantially different projections in this case would emphasize a high uncertainty about future land change processes. To quantify the difference between the chosen projections, this thesis uses three established disagreement measures and develops three novel measures which adjust the disagreement with modeling uncertainty.

- *Give guidance which helps to decide about the importance of a quantified difference:* Once a difference is quantified, the next challenge is to determine its meaning. Does a certain difference value between two projections mean that the modeled future is uncertain? This question depends on the chosen research problem. For example, a case study with a high amount of land change tends to have higher differences in two projections than a case study with few changes. The dissertation investigates this problem by suggesting a reference comparison.

5 Structure

This thesis is structured in five chapters. Subsequent to the introductory chapter, the second chapter addresses the first research objective. The second objective is investigated in chapter three and four. Chapters two, three and four contain articles which have been published or submitted to international peer-reviewed journals, and are therefore included as self-contained parts of this thesis. Each of them has its own introduction, methods, results, discussion, and conclusion sections. For that reason, repetitions are inevitable among the different chapters. The chapters are:

(I) Introduction: This chapter introduces why land change models are applied and why uncertainty is an important factor. Furthermore, the state-of-the-art, research gaps and main objectives of this thesis are given.

(II) Bayesian belief networks as a versatile method for assessing uncertainty in land change modeling: This article presents an approach based on probabilities to analyze uncertainties within the modeling process and their influence on the land change modeling results.

Krüger, C., and Lakes, T., 2014. Bayesian belief networks as a versatile method for assessing uncertainty in land-change modeling. International Journal of Geographical Information Science, 29 (1), 111–131. DOI: http://dx.doi.org/10.1080/13658816.2014.949265.

(III) Revealing uncertainties in land change modeling using probabilities: This chapter explains two newly developed uncertainty measures and their utilization for land change projections.

Krüger, C., and Lakes, T., forthcoming. Revealing uncertainties in land change modeling using probabilities. Transactions in GIS.

(IV) How similar are two land change projections?: This section deals with the assessment of uncertainty in future land change by means of a six-dimensional comparison of two or more land change projections.

Krüger, C., and Lakes, T., submitted. How similar are two land change projections?

(V) Synthesis: This chapter summarizes and discusses the findings of this dissertation in the context of the overarching research objectives. Additionally, the limitations of this study are identified and recommendations are given.

Chapter II:
Bayesian belief networks as a versatile method for assessing uncertainty in land change modeling

International Journal of Geographical Information Science
2014: 29 (1), 111–131

Carsten Krüger and Tobia Lakes

Published online: 21. August 2014.
DOI: http://dx.doi.org/10.1080/13658816.2014.949265

Abstract

Land use and land cover change modeling helps us to understand the driving factors and impacts of human-induced land changes better, and depict likely future development paths. Uncertainty associated with various steps in the modeling process substantially influences the reliability of the results, but until now it has only rarely been addressed. In this study, we explore uncertainty in land change modeling using a probabilistic approach based on Bayesian belief networks. We apply this approach to a case study of deforestation in the Brazilian Amazon and identify three modelling steps as sources of uncertainty: model structure, variable selection, and data preprocessing. For these three steps, we quantify the uncertainty and the respective impact on the outcome accuracy. The results indicate remarkable uncertainties in each of the steps. We demonstrate that a higher uncertainty in the land change modeling process does not necessarily lead to a lower accuracy of the modeling outcome. Moreover, we show that the different uncertainty sources only slightly influence the ratio between quantity disagreement and allocation disagreement for the modeling outcome. We conclude that uncertainty is inherent in land change modeling, and that future studies should address this uncertainty more explicitly to improve the robustness of modeling outcomes for science and decision-making.

1 Introduction

Rapid and spatially extensive land use and land cover change processes have dramatic impacts on the environment and society. One example is the hotspot of deforestation in the Brazilian Amazon, which imposes challenges on science and decision-making (Fearnside 2008). To gain a better understanding of the complex human-environment interaction of land systems, land use and land cover models (land change models) are increasingly used to study the rates, patterns, drivers, consequences, and alternative future development paths of land change (Verburg *et al.* 2004, Messina *et al.* 2008). However, uncertainty associated with land change modeling has up to now only rarely been studied explicitly (Chen and Pontius 2010, Wallentin and Car 2013).

While uncertainty is inherent in land change modeling, there may be differences in the sources of uncertainty, as well as in the effect of uncertainty on the reliability of land change modeling results. It is therefore of high importance to assess uncertainty, and its sources in land change modeling to adequately represent the robustness of modelling results for science and decision-making.

Uncertainty is an umbrella term that is used in different contexts without a generic definition or typology in applied research (Ascough II *et al.* 2008). Although the term is frequently used as a synonym for error, uncertainty carries more meaning (Leyk *et al.* 2005), and is associated with a lack of knowledge or confidence (Aerts *et al.* 2003, Sigel *et al.* 2010). For example, Walker *et al.* (2003) define uncertainty as “any departure from the unachievable ideal of complete determinism”. Moreover, these authors differentiate four sources of uncertainty in decision support models: context, input, parameter, and output uncertainty.

To our knowledge, a similar systematic framework does not exist for addressing uncertainty in land change modeling. However, several studies related to environmental modeling exist that address uncertainty extensively. Examples are Wallentin and Car (2013) who addressed uncertainty associated with model conceptualization, formalization, parameterization, analysis and validation for an ecological forest succession model, or Refsgaard *et al.* (2007) who formulated guidelines to deal with uncertainty in environmental modeling. Several land change modeling studies have been carried out that address a single source of uncertainty. For example, Crosetto *et al.* (2001) investigated uncertainty propagation on model outputs driven by remote-sensing input data. Pontius (2002) addressed the impacts of different spatial scales on model accuracy. Some effort has been undertaken to consider uncertainty in terms of the accuracy of land change results, and to investigate the predictive accuracy of land change models (Brown and Heuvelink 2007, Pontius and Neeti 2010). Batisani and Yarnal (2009) examined the influence of errors in input parameters on the uncertainty of urban sprawl simulations.

Different methods are available to describe and assess uncertainty, ranging from descriptive qualitative methods (Refsgaard *et al.* 2007) to complex mathematical approaches (Pebesma *et al.* 2007). Most widely established land change modeling techniques such as regression approaches address uncertainty by calculating the significance of regression coefficients (Aguiar *et al.* 2007, Arima *et al.* 2011), by plotting receiver operating characteristic (ROC) curves (Temme and Verburg 2011), and by analyzing the explained variance of the model output (Lakes *et al.* 2009, Wyman and Stein 2010). Uncertainty of future land change results can be captured by means of scenario analysis when showing the spatial consequences of different underlying assumptions (Laurance *et al.* 2001, Soares-Filho *et al.* 2006). Another widely applied approach is error propagation modeling, in which a probability distribution of errors in the model input is defined and simulates the effect on the output distribution (Heuvelink *et al.* 1989). Sensitivity analysis, which accounts for variation in the model output due to variations in an individual model input, is also frequently applied (Crosetto and Tarantola 2001). Monte Carlo simulations can be used to numerically solve complex modeling problems and to determine uncertainty in spatial predictions (van Horssen *et al.* 2002). Aldwaik and Pontius (2013) used intensity analysis to compute the minimal error related to input data, which could account for observed differences between two land change maps.

Uncertainty can be represented by a probability distribution or measures which are derived from such a distribution. Bayesian belief networks (BBNs) (Pearl 1988, Neapolitan 2004) employ a probabilistic approach that allows the researcher to more accurately estimate uncertainty compared to using only expected values (Uusitalo 2007, Aguilera *et al.* 2011). In a BBN, different predictor variables are connected in a graphical network in which relationships are commonly quantified by conditional probability tables (Pearl 1988), others use, for example, conditional Gaussian nodes (Lauritzen and Jensen 2001). The probability distribution of the child node is conditioned on the probability distributions of the given parent nodes (Pollino *et al.* 2007). New information of a parent node influences the probability of a certain state of the child node. Through the probabilistic representation in a BBN, uncertainty is integrated within the model (Stassopoulou *et al.* 1998). High uncertainty is typically modeled by a uniform distribution (Uusitalo 2007), which assigns equal probability to every possible value.

BBNs have been successfully applied in land change studies because they make it possible to combine quantitative and qualitative sources of knowledge (Thomas *et al.* 2005, Pollino *et al.* 2007), develop scenarios with varying states of influencing factors, intuitively interpret the model and its results (Aguilera *et al.* 2011), and include nonspatial and spatially explicit variables. For example, BBNs have been successfully used to analyze how characteristics of land managers affect land change in Scotland (Aalders 2008), create a decision support tool for rangeland management in Australia (Bashari *et al.* 2008), model land change decision-making in China (Sun and Müller 2013), manage willows in a river catchment in Florida (Wilkinson *et al.* 2013), and identify trade-offs between the development and conservation of landscape in the United States (Mc Closkey *et al.* 2011).

While the concept of uncertainty has received increasing attention and several studies address selected aspects of uncertainty in land change modeling, systematic approaches and quantification of uncertainty and the effects on the model's outcome in land change modeling are rare. The aim of this article is therefore to contribute to this research field by using BBNs to explicitly address and quantify uncertainty in land change modeling as captured by a case study on deforestation in the Brazilian Amazon. More explicitly we want to answer the following questions:

1) How can uncertainty in a BBN land change model be quantified?
2) What contribution do the three eminent land change modeling steps (model structure, variable selection, and data preprocessing) have on the modelling uncertainty?
3) How does the effect of uncertainty propagate throughout the modeling process and affect the accuracy of the model's outcome?

2 Materials and methods

2.1 Study area and data

We model deforestation in the states of Pará and Mato Grosso, both of which are located in the Brazilian Amazon (Figure II-1). The region contains 68 % of the total area that was deforested within the Brazilian Amazon between 1998 and 2011 (Instituto Nacional de Pesquisas Espaciais (INPE) 2013). In this study, we refer to land use and land cover change (land change) when we model deforestation. We use PRODES data ("Projeto de Monitoramento do Desmatamento na Amazônia Legal por Satélite" (Instituto Nacional de

Pesquisas Espaciais (INPE) 2013)) to assess the forest cover of a pixel in a certain time step. Subsequently, a deforested pixel can change into different land uses, such as pasture and soy production.

We focus on the deforestation process from 2002 to 2005 because these years have comparable political and economic conditions (Fearnside 2008). Based on the PRODES data, we derive a binary output. Each pixel that represents forest in 2002 and 2005 is labeled with 0 (= forest persistence, 95.73 % of all pixels). Each pixel representing forest in 2002 and non-forest in 2005 is labeled with 1 (= deforestation, 4.27 % of all pixels). Those pixels that show non-forest in 2002 remain without exception non-forested.

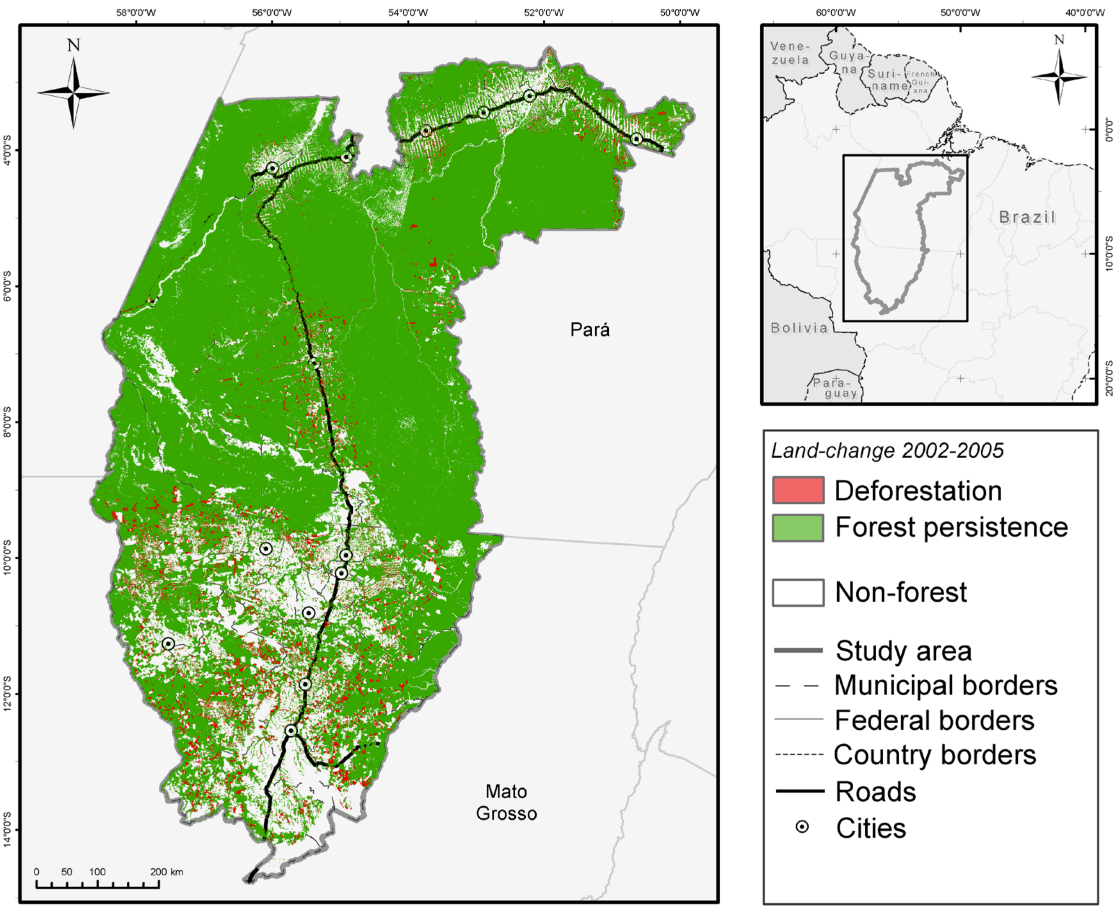


Figure II-1: Study area

Therefore, we exclude those pixels from the analysis. While there is no quantitative accuracy assessment of the PRODES data available, the “residuo” class of the data gives some evidence about misses (real change, but not detected in the data). The class quantifies the amount of deforestation, which is detected in 2012, but occurred before and cannot be assigned to any year. This class has a fraction of 0.34 % of our study area. Despite the described weakness, PRODES data have been used as a convenient data source (Barona *et al.* 2010) and are widely used in deforestation studies (Aguiar *et al.* 2012). Therefore, we assume that the data have no large systematic errors.

Based on a comprehensive literature review, we use a selection of demographic, economic, accessibility, planning, and biophysical variables as input for our deforestation models (Table II-1). To harmonize all the data to the same spatial resolution of 90 × 90 m pixel size, we apply different preprocessing steps. We disaggregate the demographic and economic data that were provided at the municipality level by assuming equal values for all pixels within one municipality. Population density is an exception and will be used later in the article to show possible consequences of data preprocessing on the modelled uncertainty. The accessibility variables are calculated using the Euclidean distance. Protected and indigenous areas are available in vector format; pixels located inside these areas are classified as 1, and pixels outside are categorized as 0. Economic areas are digitized and classified into six different classes describing the dominant economic activity based on Coy and Klingler (2008). Elevation and slope are acquired from a Shuttle Radar Topography Mission (SRTM) dataset. Original shape data of precipitation, temperature, and soil fertility are converted into the raster format with 90 × 90 m resolution.

Table II-1: Data

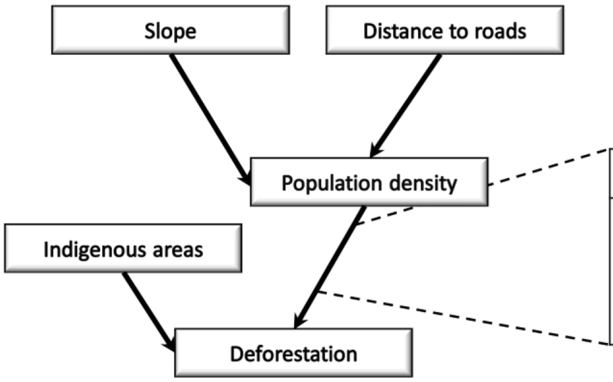
	Variable (dimension)	**Time**	**Source**
	Deforestation (0/1)	2002-2005	INPE
Demographic	Population density (inhabitants/km²)	2000	IPEA
	Share of rural population (%)	2000	IPEA
	Population growth (%)	1996-2000	IPEA
Economic	Cattle density (Cattle/km$^{2)}$	2002	IPEA
	GDP per capita (GDP/inhabitants)	2002	IPEA
	Agricultural GDP per capita (Agricultural GDP/inhabitants)	2002	IPEA
	Fraction farming/pasture (%)	1995	IBGE
	Fraction of farms > 1000 ha (%)	1995	IBGE
	Employment share agricultural sector (%)	2002	IBGE
	Employment share secondary sector (%)	2002	IBGE
Accessibility	Distance to waterways (km)	2010	MT
	Distance to major roads (km)	2007	IBGE
	Distance to Sao Paulo (km)	2007	IBGE
	Distance to airports (km)	2010	MT
	Distance to ports (km)	2010	DNIT
Specific areas	Protected areas (0/1)	2002	MMA
	Indigenous areas (0/1)	2002	MMA
	Economic areas (categories: 0-5): no special region, soy production, sawmill, smallholder, large-scale cattle ranching, former mineral extraction	2011	(Coy and Klingler 2008)
Biophysic	Elevation (m)	2007	SRTM
	Slope (%)	2007	SRTM
	Soil fertility (categories: 0-3)	2002	MMA
	Yearly average temperature (°C)	1961-1990	SISCOM IBAMA
	Yearly sum of average precipitation (mm)	1961-1990	MMA

DNIT-Departamento Nacional de Infraestrutura de Transportes; IBGE-Instituto Brasileiro de Geografia e Estatística
INPE-Instituto Nacional de Pesquisas Espacias; IPEA-Instituto de Pesquisa Econômica Aplicada
MMA-Ministerio do Meio Ambiente; MT-Ministerio do Transportes
SISCOM IBAMA-Sistema Compartilhado de Informações Ambientais do Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis; SRTM- Shuttle Radar Topography Mission

2.2 Land change modeling with BBNs

To model land change with a BBN, the target (deforestation) and the explanatory variables (see Table II-1) are represented in the directed acyclic graph as nodes connected by edges following dependencies. One variable is conditionally dependent on a second variable if there is an edge from variable 2 to variable 1. A BBN graphically represents the joint probability distribution ($P(X_1,\ldots,X_n)$; Equation II-1) over the given variables (X_i) (Pearl 1988). The probability of a state of a given child node is quantified conditioned on the states of the parent nodes.

II-1 $$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P(X_i \mid Parents(X_i))$$



		Population density		
		Low	Medium	High
Deforestation	No	65 %	70 %	20 %
	Yes	35 %	30 %	80 %

Figure II-2: Example of a graphical structure and conditional probability table of a simple BBN

The example of a simple BBN in Figure II-2 illustrates the dependency of “deforestation” on “population density”, and “indigenous areas”, that is, these nodes are parent nodes of “deforestation”, which is the child node. The conditional probability table specifies the probability of deforestation given that the population density is “low”, “medium”, or “high”.

We use deforestation from 2002 to 2005 as the target variable, and all other variables in Table II-1 as potential explanatory variables. The continuous variables are discretized into five classes with the same number of observations, whereas the nominal variables retain their original number of classes. We set up the model structure by combining expert knowledge and learning from data. Expert knowledge was derived through a systematic literature review and by an expert survey to identify relevant drivers, the confidence in those drivers, and the causal relationship between variables within the specific study area. We choose this study area and the expert survey because of a close cooperation with the ongoing project CarBioCial ("Carbon sequestration, biodiversity, and social structures in Southern Amazonia", see http://www.carbiocial.de) which focuses on sustainable land management and its effect on ecosystem services. The 15 experts who participated in the survey are from different scientific backgrounds such as land change modeling, political science, landscape ecology, and agriculture. They were asked to decide for each of the given pairs of model variables if one variable is dependent on the other. Additionally, they were asked to include a confidence value about their decision between 1 = uncertain and 3 = certain. Only those links between two variables which were identified with a high confidence by 80% of the experts (e.g., the link between population density and deforestation) are enforced in the model. Furthermore, we assume that no variable is dependent on the deforestation variable and hence constrained the structure to have no children of the deforestation node.

The remaining dependencies are learned from the data using the statistical R package "bnlearn" (Scutari 2010) with the grow–shrink algorithm (Margaritis 2003), which is a constraint-based learning algorithm. Subsequently, we apply the Bayesian parameter estimation in R to learn the conditional probability tables. To calibrate our model, 5000 random samples are selected. We choose a minimum distance of 500 m between the samples as a compromise between reducing effects of spatial correlation and having a sufficient amount of training data. We then predict deforestation spatially for the whole study area for the time period from 2002 to 2005. To assess the goodness of fit of our calibration, we randomly sample an additional set of 5000 different points. Figure II-3 summarizes the described modeling procedure.

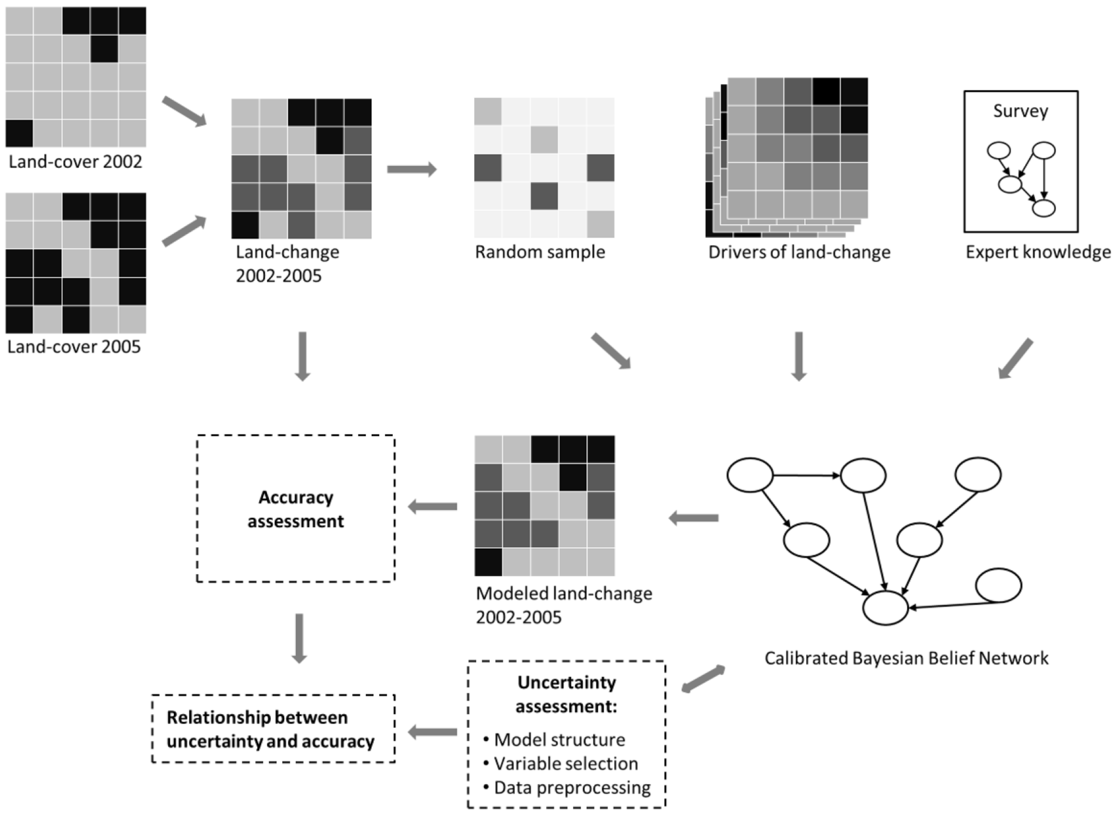


Figure II-3: Methodical procedure for uncertainty analysis with BBNs

2.3 Uncertainty analysis with BBNs

We study uncertainty in the modeling process by focusing on the three modeling steps of model structure, variable selection, and data preprocessing. Moreover, we study the effects of these uncertainties on the model's outcome accuracy, that is, on the known change. For assessing the uncertainty, we compare different model settings and calculate measures of uncertainty in each modeling step (Table II-2).

Table II-2: Uncertainty and accuracy assessment

Modeling step	**Model settings**	**Measurements for uncertainty and accuracy**
Quantifying uncertainty		
Model structure	100 different samples for data driven learning	Structure uncertainty; 95 %-confidence interval of mutual information (Pearl 1988)
Variable selection	Entire variable set vs. subset derived from stepwise BBN construction	Mutual information (Pearl 1988)
Data pre-processing	Model with original population density data vs. model with disaggregated population density data	Mutual information (Pearl 1988)
Quantifying accuracy		
Model outcome	Multiple model runs with different settings for variable selection and pre-processing simultaneously	Total error, hits, misses, false alarms, allocation disagreement quantity disagreement (Pontius and Millones 2011), ROC (Mas *et al.* 2013)

The structure of the model is used to represent the real-world problem. The *uncertainty related to the model structure* is highlighted as one (if not the) major source of uncertainty in modeling (Refsgaard *et al.* 2006, Warmink *et al.* 2010). We address this source by comparing different versions of the graphical structure of the model. We use 100 different randomly selected samples of data of the study area with each having a size of 5000 pixels. One sample is selected with replacement out of 10,000 given observations. The automated learning process described in the previous section is run on each of these 100 data samples, resulting in 100 different BBNs. Since the number of resulting possible model structures is very high (24! · 2), we focus on comparing the most crucial variables for the "deforestation" variable. Therefore, we focus on the question whether a variable is a parent node of "deforestation" and thus directly influences "deforestation" or not. No structure uncertainty is included by variables which are either never or always a parent node of "deforestation". In contrast, variables that are a parent node in some model runs and not in others lead to structure uncertainty. The fraction of the number of such uncertain nodes (N_{un}) to the total number of nodes (N_{tot}) in the calibrated BBN is used to quantify structure uncertainty (*SU*, Equation II-2). The closer the value is to 1, the higher is the proportion of uncertain nodes and the structure uncertainty, respectively.

II-2 $$SU = \frac{N_{un}}{N_{tot}}$$

Another approach to measure model structure uncertainty is to calculate the variation of the reduced uncertainty of a given variable in different model runs. High differences in reducing the uncertainty by one variable are a hint for high structure uncertainty. Uncertainty in this context is measured with the Entropy function (*H(X)*, see Equation II-3) (Pearl 1988). In the case of our binary dataset with two classes (deforestation/no deforestation) and a base of the logarithm of 2, entropy shows a maximum of 1 if both classes have the same probability of 0.5. In contrast, entropy is at the minimum of 0 if one class has a probability of 1, and the other class has a probability of 0. The advantages of using entropy for the whole probability distribution are that it is interpreted intuitively and it needs less storage capacity (van der Wel 2000). Based on the entropy, the mutual information (*MI(Y;X)*, see Equation II-4) (Pearl 1988) can be calculated. It is the potential of a variable *X* to reduce the uncertainty in the target variable *Y*. The mutual information criterion ranges from 0 (*X* and *Y* are independent) to *H(X)* (=maximum mutual information).

II-3 $$H(X) = -\sum_{x} p(x) \log_2 p(x)$$

II-4 $$MI(Y;X) = -\sum_{x}\sum_{y} p(y,x) \log(\frac{p(y,x)}{p(y)p(x)})$$

where *x* and *y* are possible values of the random variables *X* and *Y*, respectively.

The mutual information is then averaged over the 100 different BBNs based on the 100 different samples. Subsequently, we compute the 95 % confidence interval of this criterion to account for the robustness in different BBNs. If the confidence interval is narrow, the reduced uncertainty will be similar in the different model runs; this is a hint for low model structure uncertainty. However, it is possible that different structures have the same mutual information. The application of constraint-based learning algorithms such as the grow–

shrink algorithm commonly delivers independence statements which can be satisfied by different BBNs (Margaritis 2003).

We analyze *uncertainty related to variable selection* by comparing BBNs using the entire set of available variables to BBNs applying stepwise structure learning. The stepwise structure learning is applied by iteratively running the learning process: (1) Initialize the structure learning by selecting all variables. (2) Run the learning algorithm on the current set of variables. (3) If all variables have a mutual information greater than 0.01 % of the maximum mutual information, stop and report the current model. Otherwise, continue with Step 4. (4) Remove the variable with the lowest mutual information and go to 2.

We analyze these reduced networks again with sensitivity analysis and compare the results to BBNs constructed with the full set of available variables.

Uncertainty linked to data preprocessing is studied by focusing on disaggregating population data originally acquired on the municipality level with sizes of municipalities ranging up to 160,000 km^2. In many land change studies, uncertainty is assumed to be reduced by the disaggregation of spatial data; however, verifications are rare (e.g. Goerlich and Cantarino 2013). In particular, population data are frequently a limiting factor in land change studies because its availability is limited to large administrative areas (Gallego 2010). At the same time, population is stressed as an influencing variable on deforestation in this region (Laurance *et al.* 2002), as well as by our expert survey. To test the effect of disaggregation, we first exclude areas covered by water bodies and then use nighttime light data (National Geophysical Data Center 2013) to spatially allocate population density dependent on the distance to nighttime lights, assuming that population density decreases linearly with increasing distances to nighttime lights. We therefore calculate the Euclidean distance to the nearest pixel with an average light intensity value above 0. We then run and compare the BBNs with original population density data with disaggregated data by using the mutual information.

Finally, *the effects of the different sources of uncertainty on the model outcome* are analyzed by simultaneously using the different settings of the three modeling steps. We investigate whether a different amount of uncertainty substantially influences the accuracy of the model outcome. An accuracy assessment of the modeling results is possible because reference data are available. To calculate the accuracy, the modeled probabilities of "deforestation" are transferred into binary outcomes. We assume that every pixel with a probability of change below 50 % remains forest, whereas the remaining pixels change to non-forest.

Subsequently, modeled change and real change are compared using the metrics derived from the confusion matrix, which include hits (modeled as change and change in real map), misses (modeled as no change and change in real map), and false alarms (modeled as change and no change in real map). This approach allows us to differentiate the fraction of errors attributable to a wrong number of pixels assigned to a land change class (quantity disagreement (*QD*, see Equation II-5) from the fraction of errors due to incorrect spatial allocation (allocation disagreement (*AD*), see Equation II-6) (Pontius and Millones 2011). The highest possible error is equal the number of pixels in the study area, if every pixel is wrongly assigned, and the lowest possible error is 0. By dividing the errors by the number of pixels in the study area, we derive the errors in percentages.

II-5 $$QD = abs(M - FA)$$

II-6 $$AD = 2 \cdot min(M; FA)$$

where *M* is the number of misses, *FA* is the number of false alarms, *abs* is the absolute value, and *min* is the minimum value.

We then calculate the ROC curve which is an accuracy measure taking different choices of a threshold into account. It is a graphical plot which compares continuous values, such as the probabilities of land change, with true categorical values (Mas *et al.* 2013). The ROC curve allows the investigation of whether real change pixels are concentrated on modeled pixels with a relatively high probability. The relationship is quantified following Mas *et al.* (2013).

3 Results

The results of *the analysis of uncertainty associated with the model structure* are shown in Figure II-4. Apart from “population density”, which is included a priori as a parent node of “deforestation”, the variable “indigenous areas” is identified as a parent node in every model run. That means uncertainty associated with the influence of indigenous areas on deforestation is very low. Where indigenous areas exist, deforestation is certainly lower. “Distance to waterways”, “elevation”, “distance to Sao Paulo”, and “cattle density” represent parent nodes in 6-47 % of the 100 samples. This means that for those uncertain nodes we cannot clearly state if one of them directly influences the land change variable. The other variables are never identified as a direct parent so that with low uncertainty we can exclude them from explaining deforestation in our deforestation model. The fraction of uncertain nodes to all nodes as a measure of structure uncertainty is 17.39 %. The standard deviation of the mutual information supports the findings about the uncertainty of the model structure. The uncertain nodes have the highest variation. In some model runs, they have a high potential to reduce uncertainty, whereas in some runs the potential is low.

The *analysis of the uncertainty related to variable selection* shows that the uncertainty reduction as calculated by the mutual information criterion is more concentrated on fewer variables following the stepwise construction of the BBN (Figure II-5). Additionally, the stepwise learning process leads to a higher overall reduction of uncertainty compared to structure learning using the entire set of variables (23.72 % vs. 21.07 % of the maximum mutual information).

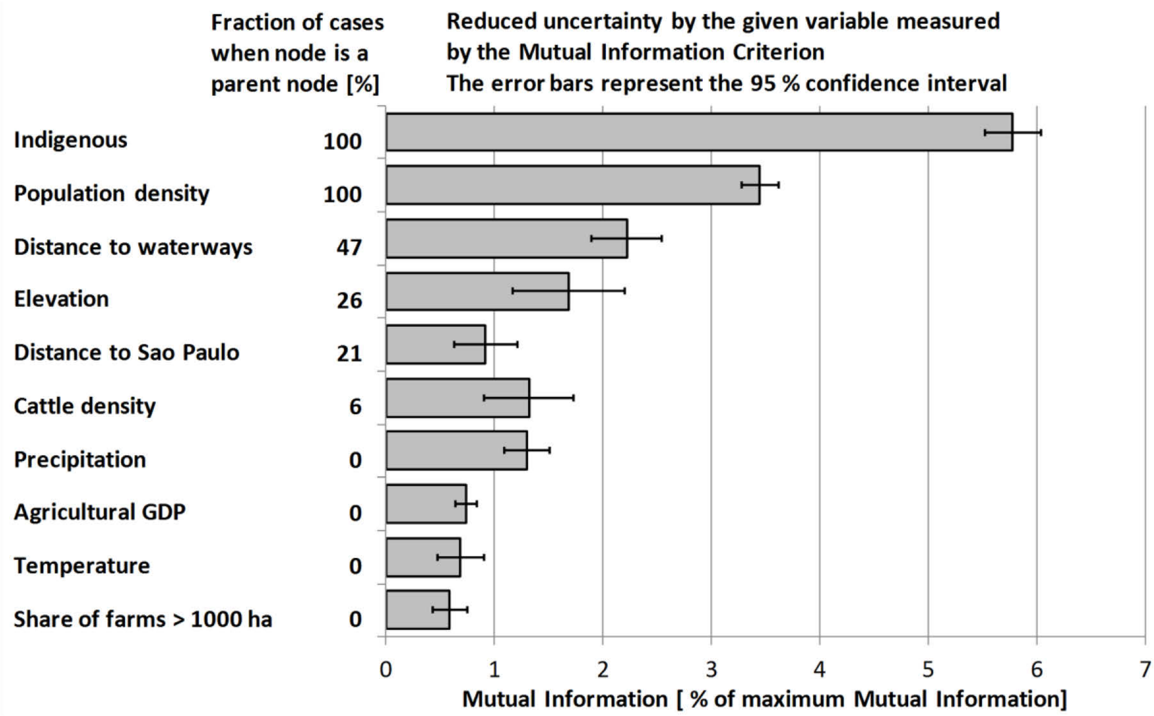


Figure II-4: Structure uncertainty for the 10 variables with the highest average mutual information

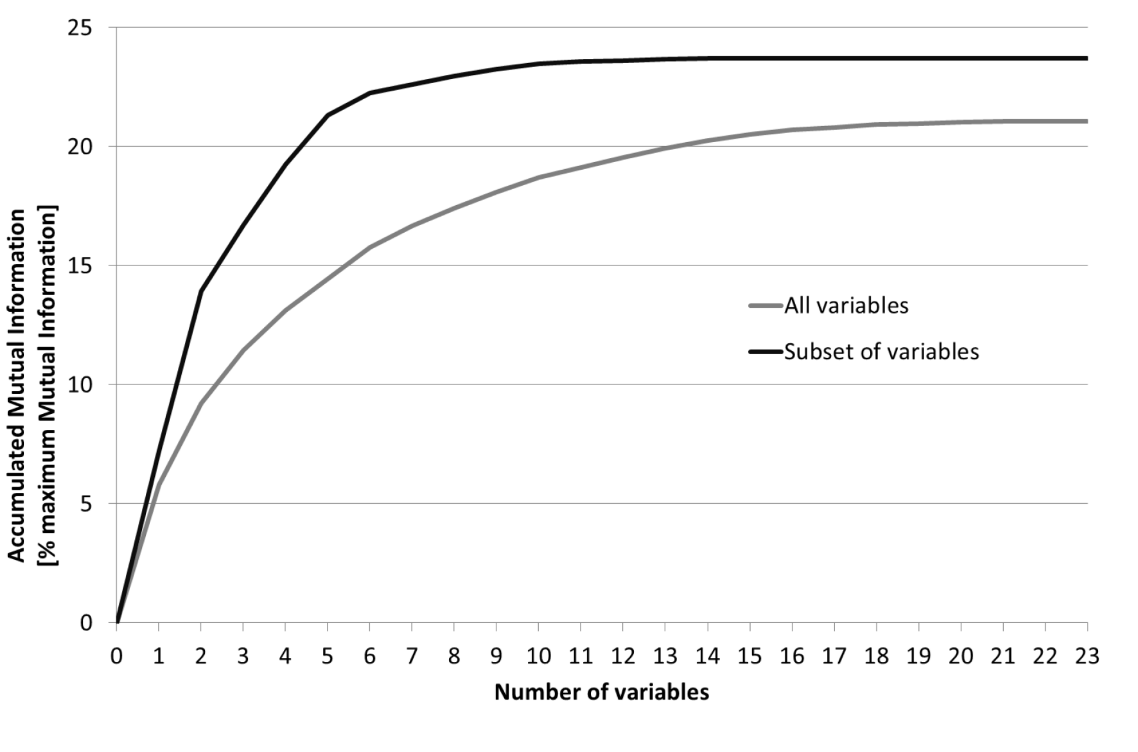


Figure II-5: Reduced uncertainty of model variables for different variable selections

The *uncertainty analysis of data preprocessing* is based on a comparison between model runs using original population density data on the municipality level and disaggregated population density data (Figure II-6). Including the disaggregated data leads to less reduction of uncertainty compared to including original data. Regarding the approach of stepwise net construction, for example, the mutual information is 0.59 % versus 2.54 % by comparing the effects of disaggregated versus original data.

The simultaneous effect of the different uncertainty sources for various model settings is presented in Figure II-7. Structure uncertainty, indicated by the SU-bar, decreases through the stepwise net construction approach and when using disaggregated population density data. At the same time, the accumulated average mutual information of all variables increases. The reduction in the overall uncertainty is higher. Therefore, this measure indicates the same tendency as the SU-bar.

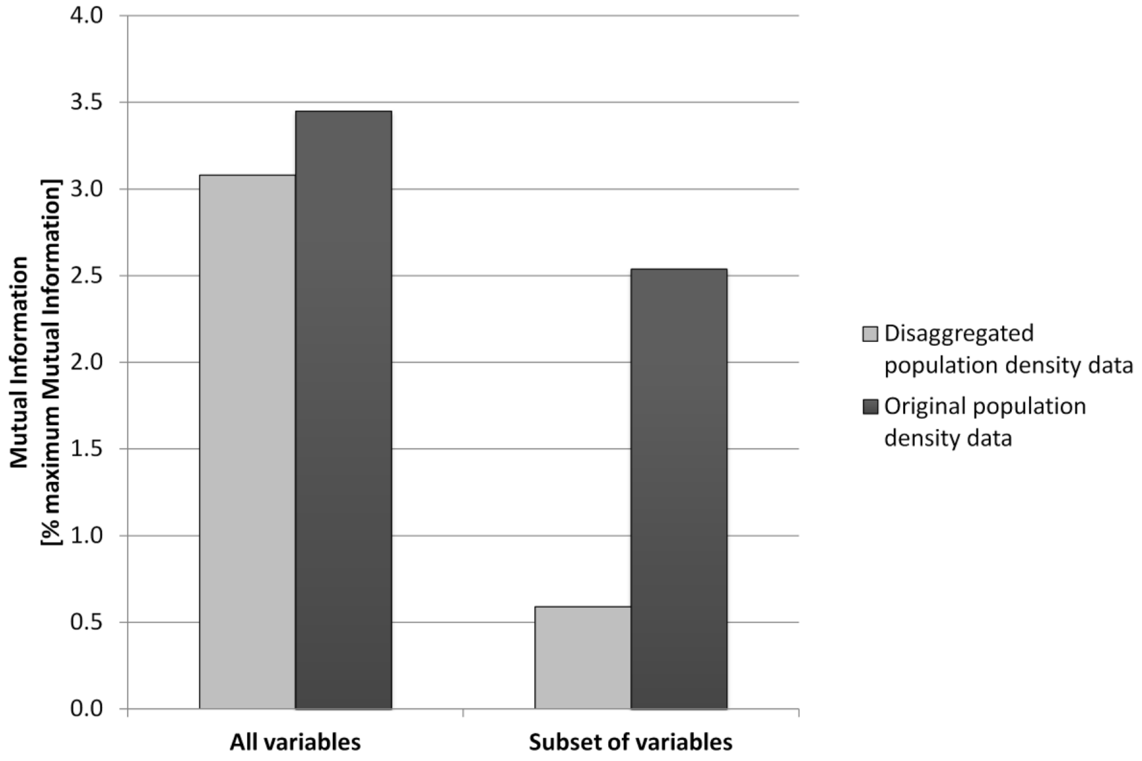


Figure II-6: Uncertainty due to data preprocessing

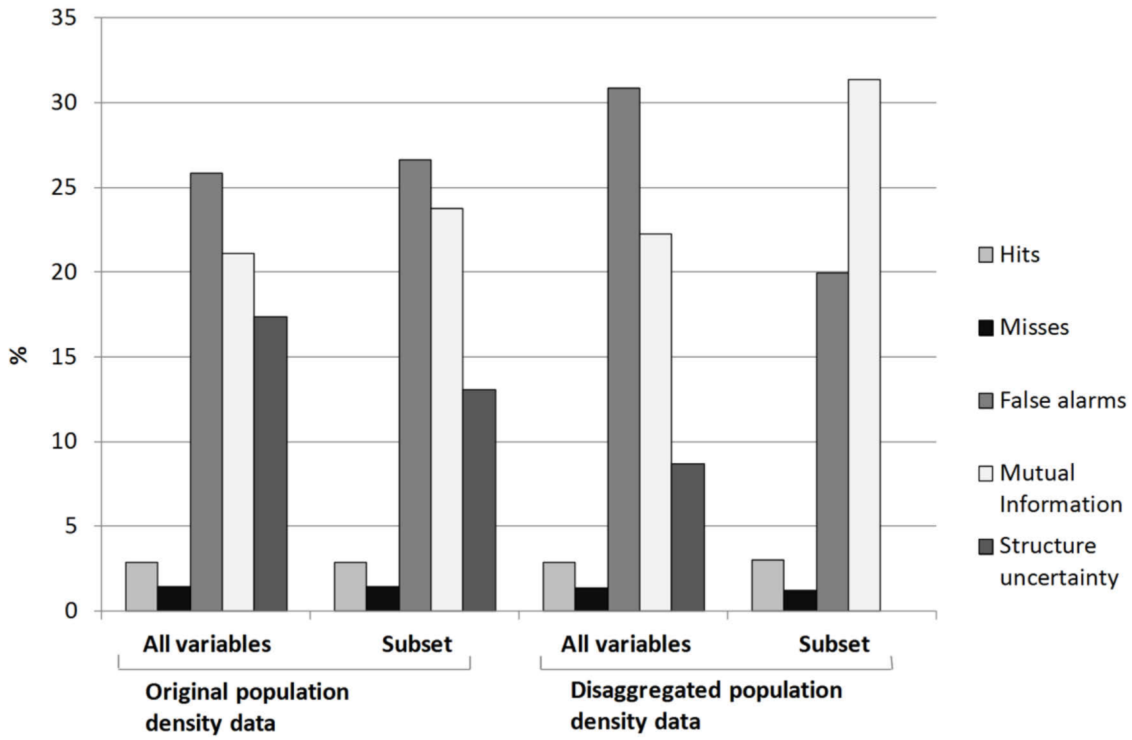


Figure II-7: Uncertainty and accuracy

Analyzing the effects of the different uncertainty sources on the *modeling outcome accuracy* shows that the ratio of quantity disagreement to allocation disagreement varies only slightly among the three sources of uncertainty (Figure II-7). The ratio is determined by the number of misses and false alarms. The number of false alarms is substantially higher, independent of the model settings. Therefore, most of the error is due to quantity disagreement. However, the total error (misses + false alarms) exhibits considerable variation. The lowest total error (21.16 %) and the lowest rate of false alarms (19.94 %) are detected by using a subset of variables, and by including disaggregated population density data. For this model setting, no structure uncertainty given by the number of uncertain nodes is calculated. Additionally, the highest reduction of uncertainty given by the mean accumulated mutual information is measured (31.35 %). The fraction of hits has similar values of between 2.85 % and 3.05 % dependent on the different model settings.

Since we assume that every pixel with a probability above 50 % is predicted as deforestation pixel, a total of 23.01 % of the pixels are modeled as deforestation, whereas 4.29 % are classified as deforested in the reference data. The difference between modelled deforestation and real deforestation leads to a high fraction of quantity disagreement. Figure II-8 points out that the areas with probabilities above 50 % are larger in comparison to the real change in Figure II-1. At the same time, the map shows a good spatial match between modeled and real deforestation. The ROC curve in Figure II-9 emphasizes the relationship that real change pixels are concentrated on modeled pixels with a high probability. The bold marked point symbolizes the 12.5 % fraction of all pixels with the highest probability. The true positive rate related to this point is 0.33 in comparison to the false positive rate of 0.11. The point is above the diagonal representing a random model. The detected errors of the land change model are slightly spatially correlated (Moran’s I of 0.21).

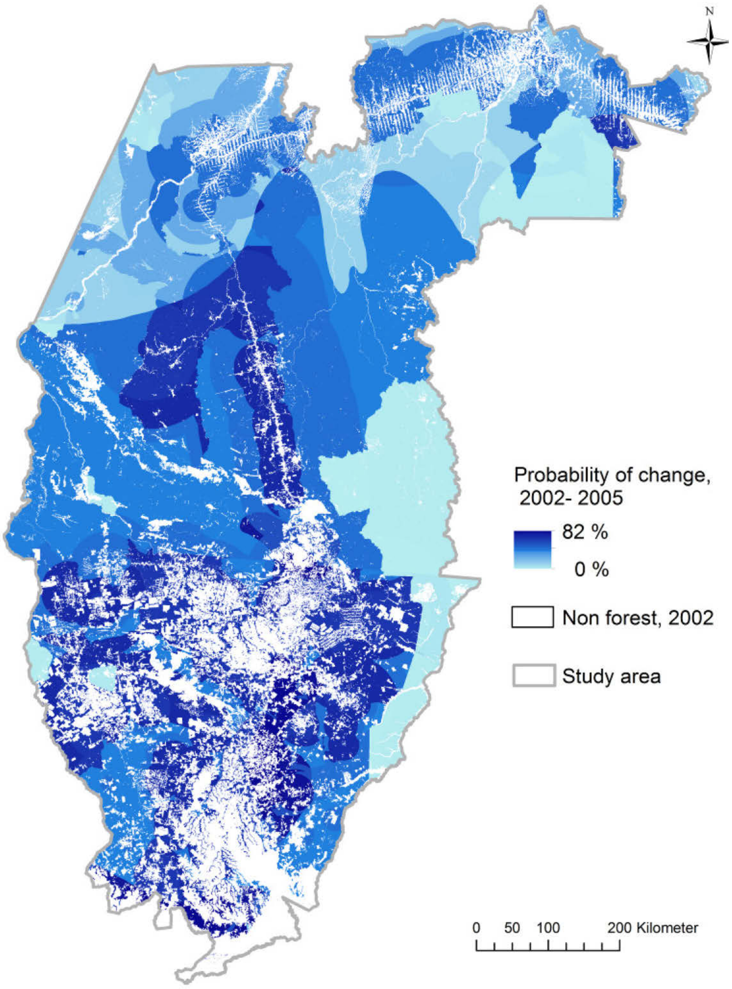


Figure II-8: Spatial distribution of modeled probabilities

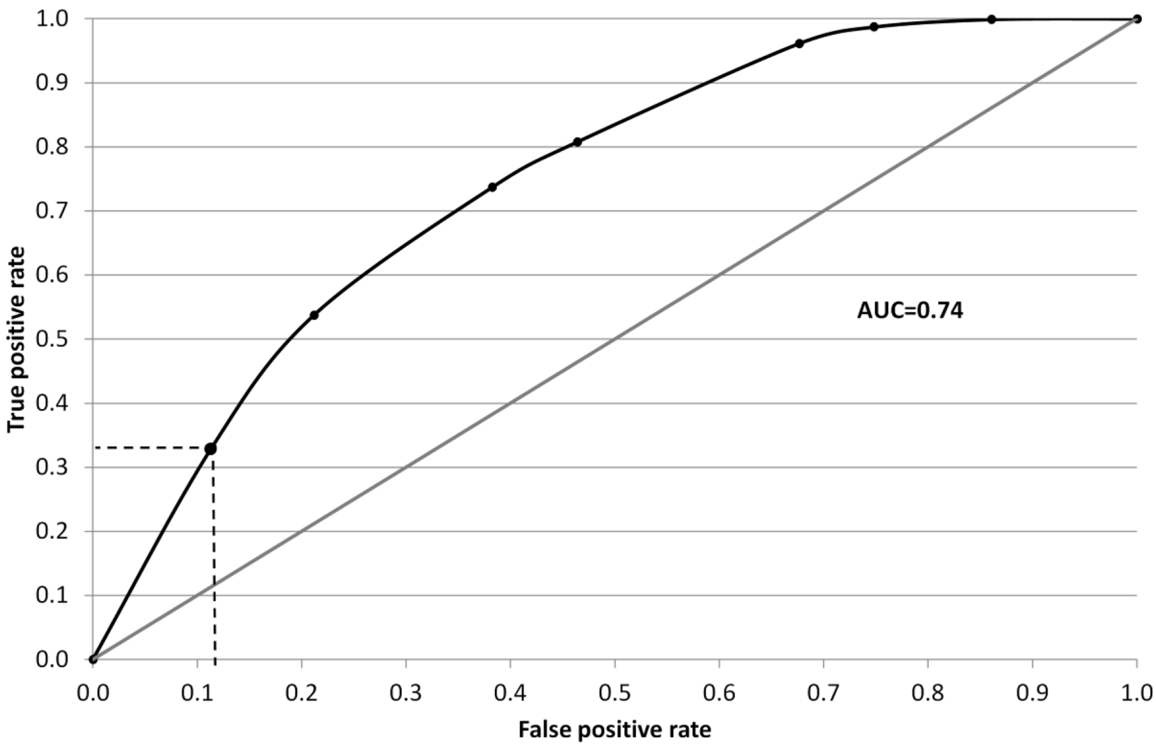


Figure II-9: Receiver operating characteristic of modeled deforestation from 2002 to 2005; AUC-area under the curve

4 Discussion

This work presents a systematic approach for addressing and quantifying uncertainty in land change modeling using BBNs. We select the three essential modeling steps of model structure definition, variable selection, and data preprocessing. We therefore expand earlier studies that focus on a single step (Crosetto *et al.* 2001, Pontius 2002). Additionally, we analyze the effect of uncertainty on outcome accuracy.

We evaluate BBNs as a suitable approach to investigate uncertainties in land change modeling (Research Question 1). Uncertainty from data and experts can be integrated (Bromley 2005), which is one major reason why they are increasingly applied in environmental modeling (Aguilera *et al.* 2011). The probabilistic relationships between variables are intuitively interpretable and express a degree of uncertainty. The probability distribution of the target node conditioned on the states of the input nodes is additionally calculated. Therefore, certain and uncertain land change in spatial areas can be identified. In comparison to the widely applied ordinary least squares regression analysis, BBNs can include uncertain prior information (Fenton and Neil 2012). Thus, BBNs are suitable for dealing with incomplete or missing data by using this prior information, which itself reflects the state of knowledge before any analysis is done (Uusitalo 2007). However, one challenge of BBNs may be large conditional probability tables through a variety of parent variables of a certain node. One way to avoid this effect was suggested by following the guidelines of Marcot *et al.* (2006) or by Boutilier *et al.* (1996) who use context-specific independence. Once a BBN is compiled, it delivers fast responses to scenario or "what-if" analysis (Uusitalo 2007). We can directly explore the consequences for the target node once an observed node is fixed at a certain state. Identifying the uncertainty contribution of different sources is therefore a straightforward task. The uncertainty analysis with the BBN approach is done in a spatial context. Relationships in complex human-environment interactions are not necessarily constant in an entire study area (Fotheringham *et al.* 2002). Our study addresses spatially varying relationships by using nodes representing geographic subareas. These nodes are "protected areas", "indigenous areas", and "economic areas". Spatial interactions, however, have up to now hardly been addressed by BBNs (Duespohl *et al.* 2012). The degree of autocorrelation of spatial data is one indicator of these interactions. The calculated Moran's I of the modeled errors points out a slight amount of spatial correlation. One possibility to further address this issue could be by means of graphical models with undirected arcs; however, exact inference is not possible in that case (Laskey *et al.* 2010).

To answer our second research question (what contribution different modeling steps have on the modeling uncertainty), we focus on the three eminent land change modelling steps of model structure, variable selection, and data preprocessing. We suggest one approach to quantify structure uncertainty by means of BBNs and uncertain nodes. Our findings show that the amount of structure uncertainty is strongly dependent on the different model settings. However, the results additionally imply that most of the model variables are either always or never a parent node of land change in different model runs, which means that we assign no structure uncertainty regarding those variables. Instead, uncertainty is associated with the remaining variables. This is also reflected in the wider confidence interval of the mutual information criterion, because these variables are part of the model structure which may or may not influence land change. These variables are uncertain nodes. Other nodes that are not parent nodes could also include structure uncertainty because they can influence land change when no information about the parent nodes is given. Furthermore, we consider the effect of changing the states of one variable when we calculate the mutual information. Varying different variables simultaneously could give additional insights. The confidence interval of the mutual information should serve as additional information describing structure uncertainty. Structure uncertainty is still possible when receiving the same mutual information because of different BBNs which can satisfy the same independence constrains. Nevertheless, we think that the chosen measures are a good compromise between the complexity of measuring different structures and the substance of the measures. From our study, we suggest to consider the two measures complementarily. Through the probabilistic representation of the model structure, we are able to expand earlier studies that qualitatively describe and stress uncertainty associated with model structure (Refsgaard *et al.* 2006) and can offer quantitative information for better dealing with uncertainty and risk (Uusitalo 2007).

We compare models with the full set of available variables with models based on a reduced set of variables derived by a stepwise model construction to analyze input uncertainty related to variable selection. Results indicate that the reduction in uncertainty is higher for the more important variables and marginally lower for the less important variables if we consider the models with a reduced subset. The reduction of uncertainty measured by the mutual information criterion is more unequally distributed in favor of about one-third of the variables. The uncertainty related to all variables is difficult to assign to a certain factor if all potential variables are included in the model. More variables do not necessarily lead to a decrease of uncertainty. These results also support the finding of Walker *et al.* (2003) that different sources of uncertainty are interrelated. While additional model variables may reduce uncertainty through additional information, they may also increase model structure uncertainty (Ascough II *et al.* 2008).

We investigate input uncertainty related to the data preprocessing using the population density data as an exemplar. The disaggregated data lead to a lower reduction of uncertainty in comparison with the population density data, assuming equal values within the same municipality. We expect the disaggregated alternative to be a superior predictor if the local population in the direct neighborhood is predominantly responsible for deforestation. One potential reason for this counterintuitive result arises, when the population in the entire region indirectly influences deforestation. In that case, the original data might lead to better results. This indicates that a more complex model does not necessarily reduce uncertainty. We cannot rule out the possibility that disaggregating other drivers or taking a nonlinear decay function for population density may lead to an additional reduction of uncertainty. However, we can assess the aim to quantify the importance of a preprocessing step.

Our results show that uncertainty propagates throughout the modeling process and substantially influences the outcome error (Research Question 3). Additionally, the different uncertainty sources influence each other. Structure uncertainty is reduced by using fewer variables and disaggregated population density data, while the sum of reduced uncertainty increases. Less structure uncertainty also leads to less input uncertainty in our case study. In other words, parsimonious models may often be the preferred choice when the aim is to reduce outcome uncertainty in the land change model. Regarding the real implications of these uncertainty sources on the model's accuracy, we see the lowest total error when the structure and input uncertainty are the lowest. We can also show how the different uncertainty sources influence the composition of the error. In this study, independent of the level of uncertainty, the difference between misses and false alarms remains high. The different models predict substantially more deforestation than accounted for in the reference map. Therefore, quantity disagreement is the dominant contributor to the overall error. We use a probability of 50 % as the threshold to decide if a pixel will be assigned into the class "no change" or "change". We use a threshold to enable the application of the accuracy measures quantity disagreement and allocation disagreement. We are aware of the problem that another threshold could lead to other accuracy results. Therefore, future work will address how to use probabilities instead of a crisp classification to address errors due to an incorrect quantity and spatial allocation.

Our model is predominately developed to predict land change spatially. Therefore, all uncertainty investigations are done for this calibration step. Future work will address uncertainty in temporal land change predictions. In our study, however, we are able to show that different steps in the modeling process potentially influence the degree of uncertainty and the effects on the model outcome. Early steps such as formulating the model's structure broadly define the frame of the land change model, whereas subsequent steps refine the model. Therefore, early steps have a greater potential to include uncertainty and to influence the modeling results. Moreover, we confirm the findings of von Krayer von Krauss *et al.* (2006), who show that each modeling step depends on the steps preceding it, and that such a dependency may include unexpected complexities. Quantifying the uncertainty effects may be a helpful initial step toward developing adequate methods to deal with these dependencies.

The quantification of uncertainty that is associated with our land change model and its results has shown to add important information for interpreting the applicability and reliability of the resulting maps. For the deforestation process in the Brazilian Amazon, we identified the variable "indigenous areas" which directly influences deforestation without any structure uncertainty. Our findings hence confirm earlier studies of de Espindola *et al.* (2012) and Soares-Filho *et al.* (2006). Furthermore, “population density”, which is connected as a parent node of deforestation by means of expert knowledge, is identified as an important driver of deforestation associated with low uncertainty. “Distance to water ways” and “distance to Sao Paulo” are additional crucial drivers of deforestation in a large share of model runs. They represent accessibility variables which indicate that easily reachable areas are prone to deforestation. However, we are not completely certain about their influence on deforestation, because these variables are uncertain nodes. A clearer picture could motivate political decision-makers to enforce the constitution of protected areas in these locations to avoid further deforestation. Our results, hence, confirm the ambiguity of the effect of distance to infrastructure on deforestation that was also highlighted by the findings of de Espindola *et al.* (2012). They showed as well that easily accessible areas are more likely to be deforested (“distance to roads” is a major driver); however, that “distance to rivers” is less relevant. In addition, our study identified several variables which have a low influence on deforestation with a low structure uncertainty as well, such as “distance to roads”, “GDP per capita”, and “Slope”. “Distance to roads” is the variable of this set, which is mentioned as an important driver of deforestation in other studies (Aguiar *et al.* 2007). However, the interpretation of the association between the distance to roads and deforestation remains difficult because of the potential problem of endogeneity.

Concerning the reliability of the overall model, we are relatively certain about the location which is close to previously deforested areas, but uncertain about the correct quantity of land change. In other words, we know which pixels are highly susceptible for deforestation, but we do not know the real amount which is highly dependent on global variables. The derived information about the uncertainty associated with the model of deforestation in the Brazilian Amazon adds important information for decision-makers and stakeholders. On the one hand, we know with a high certainty which areas are prone to deforestation, and therefore we can allocate ecological threats. On the other hand, we can evaluate which drivers are important for the deforestation process and show a low uncertainty. Incentives which try to reduce deforestation should especially consider these drivers. In contrast, potential drivers with a high uncertainty should be further analyzed. Moreover, a high uncertainty in the model outcome that includes an analysis of the uncertainty associated with the different modeling steps helps to effectively improve the land change model and therefore its results. In a next step, we can use the calibrated model with the known values of associated uncertainty to predict different future development paths of deforestation by including different delineations of indigenous areas, for example.

5 Conclusions

In this article, we show that BBNs are a versatile method both for modeling land change and for explicitly addressing different sources of uncertainty in land change modeling and their effects on the accuracy of the model's outcome. Moreover, we suggest a set of measures to quantify uncertainty by using the mutual information criterion and the number of uncertain nodes in a BBN. Additionally, accuracy of the modeling outcome is assessed by different indices. Our BBN-based uncertainty framework is a beneficial contribution for dealing with uncertainty in the land change modeling community, where systematic concepts are rare. Our study does not end at the quantification of uncertainty; we also investigate whether this uncertainty has real implications on the measurable accuracy of the modeling outcome. We focus our analysis on three different sources of uncertainty, which should not imply that no other sources can potentially have a substantial influence. We choose these three sources to cover the most frequently mentioned challenges in previous land change modeling studies. Our land change model analyzes deforestation in the Brazilian Amazon. Here, uncertainty is a substantial factor, stemming both from unknown and informal processes that lead to deforestation, as well as a data pool that is partially limited regarding its accuracy and spatiotemporal resolution. We suggest further testing of our approach in other case studies, since uncertainty is omnipresent in land change modeling.

Acknowledgements

We thank the participants of the land change modeling workshop of CarBioCial ("Carbon sequestration, biodiversity and social structures in Southern Amazonia", see http://www.carbiocial.de) for their willingness to share their expert knowledge. The authors would also like to thank Dennis Funke for data preparation work, and particularly Florian Gollnow, Ulf Leser, and Daniel Müller for their helpful comments and fruitful discussions that helped to improve this manuscript. We additionally thank the anonymous reviewers which helped us improving an earlier version of the article by giving a variety of beneficial remarks and suggestions.

Funding

The presented work is funded by Deutsche Forschungsgemeinschaft (German Research Foundation) [GRK 1324/1, METRIK].

Chapter III:
Revealing Uncertainties in Land Change Modeling Using Probabilities

Transactions in GIS, forthcoming

Carsten Krüger and Tobia Lakes

Abstract

Land change models are frequently used to analyze current land change processes and possible future developments. However, the outcome of such models is accompanied by uncertainties that have to be taken into account in order to address their reliability for science and decision-making. While a range of approaches exist that quantify the disagreement of land change maps, the quantification of uncertainty remains a major challenge. The aim of this paper is therefore to reveal uncertainties in land change modeling by developing two measures: quantity uncertainty and allocation uncertainty. We choose a Bayesian Belief Network modeling approach for deforestation in Brazil to develop and apply the two measures to the resulting probability surface. Quantity uncertainty describes the uncertainty about the correct number of cells in a land change map assigned to different land change categories and allocation uncertainty expresses the uncertainty about the correct spatial placement of a cell in the land change map. Thus, uncertainty can be quantified even in those cases where no reference data exist. Informing about uncertainty in probabilistic outcomes may be an important asset when land change projections are being used in science and decision-making and moreover, they may also be further evaluated for other spatial applications.

1 Introduction

Most of the Earth surface has been altered by human activities such as agricultural expansion, deforestation or urbanization (Foley *et al.* 2005). Scientists and decision-makers strive to understand the underlying processes and impacts of these land use and land cover change (land change) processes. Therefore, different techniques (Pontius *et al.* 2008; Kim 2010; Brown *et al.* 2013; Tayyebi, A. *et al.* 2014) of land change modeling have been developed that aim to identify important drivers of land change (e.g. Lakes *et al.* 2009; Müller *et al.* 2012; Gollnow and Lakes 2014), simulate likely future developments based on scenarios (e.g. Verburg *et al.* 2002; Asselen and Verburg 2013) and provide insights for decision-making by informing about policy impacts (e.g. Schaldach *et al.* 2013). To avoid misleading interpretations, information about the reliability of land change model outcomes is of utmost interest (Tayyebi, A.H. *et al.* 2014).

One common outcome of a land change model is a map which indicates areas where the process of change is likely. A specific value is assigned to every pixel in such an outcome map which frequently represents a propensity or probability of change (e.g. Mas *et al.* 2014). Propensity values express the chance of land change at a specific location in the grid in relation to other cells; they do not inform about the quantity of change. In comparison, the quantity of change is implied in the probabilities. Most often the continuous outcome values are assigned into different classes, e.g. “change” and “persistence”.

To assess the accuracy of such modeling outcomes, a variety of measures have been developed to determine the agreement and disagreement of modeled land change by comparing expected with observed land change. Most of these measures are based on the confusion matrix, which compares modeled land change with land change in the reference data (see for example Olofsson *et al.* 2014 for detailed recommendations about using the confusion matrix to assess the accuracy of remote-sensing derived land change maps). The most commonly used measure derived from this matrix is the overall accuracy, i.e. the ratio between all correctly classified pixels and the total number of pixels. The user’s and producer’s accuracy are also widely established. The user’s accuracy gives the ratio between the correctly classified pixels of one class and the total number of classified pixels in this class. In comparison, the producer’s accuracy is the ratio between the correctly classified pixels in a specific class and the total number of pixels in this class in the reference data. An additionally widely applied measure is the Kappa coefficient (Cohen 1960), which quantifies the accuracy in relation to the expected agreement. The Fuzzy Kappa (Hagen 2003) extends this concept by considering spatial fuzziness and fuzziness between different land change categories. An enhancement of the Fuzzy Kappa coefficient has been developed which includes the consideration of spatial autocorrelation (Hagen-Zanker 2009). Pontius (2000) developed two measures based on the original Kappa coefficient which represent an early attempt to differentiate between the different accuracy components of the correct location and the correct quantity of land change pixels. In a later study, Pontius *et al.* (2008) suggested similar separations of the overall disagreement. Pontius and Millones (2011) claimed that these measures should be used instead of the Kappa, mainly because Kappa uses a random baseline, which is not a realistic one in the majority of case studies. Van Vliet *et al.* (2011, 2013) addressed this issue by the correction of the Kappa for the size of the land change classes. The receiver operating characteristic (ROC) is another measure which uses a random baseline as a reference comparison; however, it takes different possible class sizes into account. The graphical plot compares continuous values, such as the propensity of land change, with true categorical values (e.g. Mas *et al.* 2013). The Y-axis of the plot gives the rate of correctly classified pixels (true positives) for various threshold levels, whereas the X-axis illustrates the rate of false alarms (false positives). To summarize the ROC, the area under the curve (AUC) is frequently calculated (e.g. Lakes *et al.* 2009). Lobo *et al.* (2008) summarized some problems that occur when only this value is given and in a recent article Pontius and Parmentier (2014) developed some additional recommendations for the interpretation of the ROC curve. Apart from the separation of the disagreement into quantity and spatial components, some accuracy studies specifically address the spatial variation of accuracies. Foody (2005) measured accuracies of different subregions and different thematic classes. Other studies used the Geographically Weighted Regression approach to present a spatial accuracy distribution (Comber *et al.* 2012; Comber 2013). They estimated a spatially explicit accuracy value as a function of the two-dimensional space.

The limitation of each of the measures outlined above is that they require reference data representing the real land change. Therefore, they cannot inform about accuracies of land change modeling outcomes for future time steps when no reference data is available. However, to be able to assess the goodness of the modeling results, the concept of uncertainty can be explored. Uncertainty represents “any departure from the unachievable ideal of complete determinism” (Walker *et al.* 2003). This implies that a true value exists; however, the value is not known or is uncertain. Two general approaches are frequently applied to describe the true value under uncertainty: either a delimitation of the existing space likely containing the true value, or the probability distribution (e.g. Laskey *et al.* 2010; Mc Closkey *et al.* 2011) assigning a probability of being the true value to every possible value, is given.

The first approach can be implemented by giving the standard deviation, confidence interval, range, coefficient of variation or significance level (Gopal 2009; Malizia 2013). For the second approach, the probability distribution can be compressed to the entropy measure (Shannon and Weaver 1949), which is an uncertainty measure originating from information theory (van der Wel 2000). A variable with an equal probability of all states has a high uncertainty. In the case of a binary variable, a probability of 0.5 for both states leads to the maximum uncertainty represented by the entropy measure. The quadratic score (Glasziou and Hilden 1989) is an uncertainty measure similar to the entropy measure (van der Wel 2000). Another alternative is to use the whole probability distribution for the uncertainty analysis. For example, Sangermano *et al.* (2012) compared the distribution of continuous probability values in areas of change and areas of persistence.

Several land change studies addressed and quantified the accuracy of the modeling outcomes for the processes under investigation, such as deforestation, the land change process considered in this study. Kim (2010) compared different land change models in terms of their usability in predicting deforestation processes in a part of Bolivia. The author used accuracy measures based on the confusion matrix. Sangermano *et al.* (2012) used the AUC value to show the importance of different explanatory variables for deforestation in the same country. Mas *et al.* (2014) reviewed different land change models. They analyzed the outcome of four widely-used modeling software packages for a virtual deforestation case study. Similar to this study, they included the continuous model output as well as the classified land change maps in their comparison. Pontius *et al.* (2007) evaluated the output of a land change model in a part of the Amazon basin by means of spatial and quantity disagreement measures. This case study is similar to the area used in this article. Sloan and Pelletier (2012) performed a similar division of the overall disagreement for a national deforestation study in Panama.

As outlined above, there are several measures available to characterize the agreement and disagreement in terms of spatial and quantity components. However, to our knowledge, no similar study has been undertaken to separate the uncertainty based on probabilistic data, which is an important challenge in spatially explicit land change modeling. The aim of this article is therefore to develop uncertainty measures for land change probability surfaces which differentiate between uncertainty of the correct quantity of a land change category and uncertainty of the spatial allocation of such a category. Moreover, we analyze the relationship in terms of a bivariate correlation between the proposed uncertainty and established disagreement measures, i.e. is high uncertainty associated with high disagreement? An existing relationship allows reasoning about the future reliability of land change models when only the calculated uncertainty exists. We develop and test our measures for deforestation modeling using Bayesian Belief Networks (BBNs) in Brazil.

2 Data and Methods

We first introduce our case study and then the selected land change modeling approach of Bayesian Belief Networks. Since many land change modeling techniques produce propensities of change instead probabilities as an outcome, we convert these propensities into probabilities. For the resulting probability surfaces, we develop measures to describe the quantity uncertainty and the allocation uncertainty. Finally, we compare these values with established measures of disagreement (Pontius and Millones 2011). The methods described in this article are applicable for land change modeling techniques that return a propensity or probability surface of binary land change, i.e. “change” and “persistence”.

2.1 Study area

We develop and test our approach using a deforestation case study the Brazilian Amazon which is a hotspot of deforestation (de Espindola *et al.* 2012). In 2006, the accumulated deforested area in Brazil reached 17.8 % of the original forested area (Fearnside 2008). The study region is located in the two Brazilian states Pará and Mato Grosso and comprises 51 municipalities (Figure III-1). The region covers an area of 668,393 km^2 with forest as the main land cover type in 2002 (78 %). Between 2002 and 2005, 4.14 % of the forested land was deforested while no forest gain was detected in the PRODES input data (Instituto Nacional de Pesquisas Espaciais (INPE) 2013). Based on a thorough literature review, we select a set of 23 explanatory variables from socioeconomic, biophysical, political, and accessibility measures known to influence deforestation (Laurance *et al.* 2002; Coy and Klingler 2008; Araujo *et al.* 2009; Walker *et al.* 2009; Celentano *et al.* 2012; de Espindola *et al.* 2012) (Table II-1). Since we choose this case study experiment to illustrate the application of the uncertainty measures, we do not explain the selection of data in detail but instead refer to (Krüger and Lakes 2014).

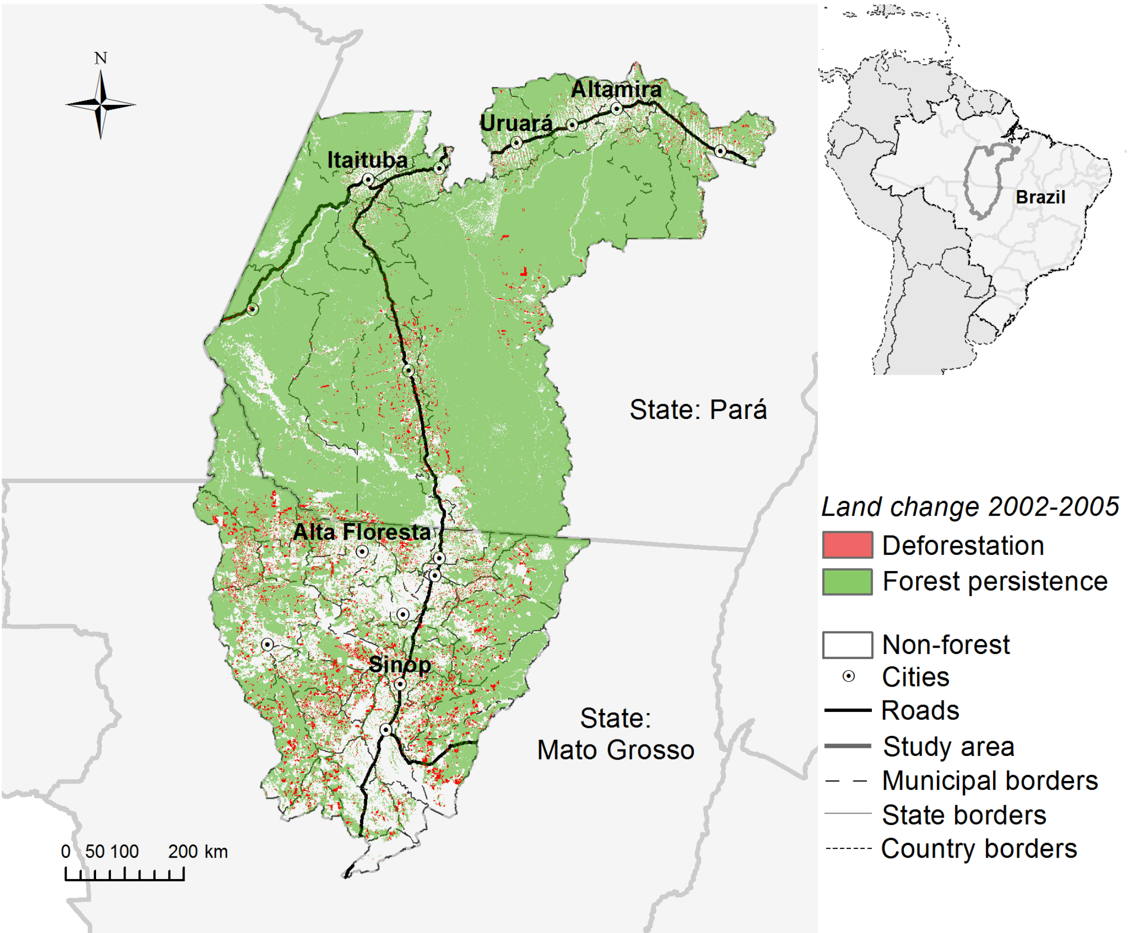


Figure III-1: Study area

2.2 Land Change Modeling with Bayesian Belief Networks

We use a Bayesian Belief Network (Pearl 1988) to model deforestation in the case study region. A BBN is a graphical representation of conditional dependencies between nodes or, in other words, model variables which are quantified by conditional probability tables. It represents the joint probability distribution ($P(X_1,\ldots,X_n)$) of the given model variables (X_i) (Equation III-1). The probabilities of the two states "change" and "persistence" of the land change variable are based on the states on the parent nodes, which represent explanatory variables.

III-1 $$P(X_1,\ldots,X_n) = \prod_{i=1}^{n} P(X_i \mid Parents(X_i))$$

To develop a BBN, two major steps have to be performed: 1) defining the structure, i.e. where and in which direction do links between the given variables exist, and 2) defining the shape of the probability distribution. Both steps can be completed through the use of expert knowledge or data driven. BBNs have already been successfully applied in land change studies (e.g. Peter *et al.* 2009; Kocabas and Dragicevic 2013, Sun and Müller 2013). Advantages of using BBNs are, e.g., the representation of uncertainty by means of probabilities, the possible reasoning from causes to effects or from effects to causes, the intuitive graphical representation of dependencies, the possible integration of expert knowledge and empirical data, the possibility to deal with missing or incomplete data, and the incorporation of non-linear relationships and correlations between explanatory variables (Aguilera *et al.* 2011; Kocabas and Dragicevic 2013).

We use the R package bnlearn (Scutari 2010) to model land change. For the calibration of our model we select a stratified sample of 10,000 pixels (90x90m resolution) of the target dataset (binary: deforestation yes/no between 2002-2005) and the explanatory dataset (23 variables). After completing a set of test runs with different sample sizes, we chose 10,000 pixels as a compromise between a computational efficiency and a sufficient amount of training data. Special economic areas are included as an explanatory variable (Coy and Klingler 2008: soy production, small scale farming, cattle ranching, mining or wood processing as dominant economic sectors) and reflect the spatial heterogeneity which is often a challenge in land change science (Fotheringham *et al.* 2002). The calibrated BBN is applied to model land change for the whole study area for the same time period 2002-2005, except for the 10,000 pixels, which are used for the calibration. The outcome of the model gives probabilities which indicate where deforestation in the study region is more likely. By means of the real change data, the amount of land change pixels can be specified. This is necessary to obtain a hard classification of the land change model output (as described below). The modeling strategy is done for the interpolation of land change in space. However, the uncertainty concepts are applicable in past, present and future modeling time steps. Additionally, more complex models are imaginable. However, we try to keep the modeling approach reasonable and easy to follow. This approach supports the focus on how to measure uncertainty.

2.3 From Propensities to Probabilities

By means of the BBN-model, we derive propensities for the different transitions of the land change variable. We aim to use the probability of the modeled land change in each cell to calculate the uncertainty about the real land change. A value of 0 implies absolute certainty that no land change will occur, 1 implies absolute certainty that land change will occur and 0.5 represents the maximum uncertainty about land change. The higher the difference to 0.5 is, the higher the certainty about the real land change class is.

Classifying all pixels with a propensity greater or equal 0.5 into the first class and the other pixels into the second class does not necessarily lead to a realistic proportion of these classes in the final land change map. If events are rare, it is possible that no pixel reaches a value above 0.5 even though some events exist (Marcot *et al.* 2006).

Propensity values express only the chance of land change at the specific location in the grid in relation to other cells and do not inform about the quantity of change. A propensity above 0.5 does not mean that “change” is more probable than “persistence”. To transfer a surface of continuous propensity values into a crisp classification of land change transitions we need to know the amount of change. Most frequently, the amount of change is defined externally from the model and then those pixels with the highest propensity values are classified as change pixels (Verburg and Veldkamp 2004; Müller *et al.* 2012). In our case, we derive the probabilities by means of the quantity of land change detected in the reference data.

We transfer the propensities into probabilities following Pontius and Batchu (2003). The logic behind these calculations is as follows: The threshold will be reduced to 0.5 if the original propensity threshold is above 0.5. The range of propensity values above 0.5 is stretched to fit the probabilities of 0.5 to 1. The range of propensity values below 0.5 is adjusted to the probabilities of 0 to 0.5. Every propensity value above/below the threshold is multiplied with the same factor to stretch/adjust the range of propensity values. The calculations are vice versa if the original propensity threshold is below 0.5.

Figure III-2 shows one example where we assume a quantity of 20 % of change. Point P_1 depicts that 80 % of all pixels have a propensity up to 0.8. Therefore, 20 % of the pixels should have a probability above 0.5 and 80 % of the pixels should have a probability below 0.5. After applying the mathematical transformation, point P_2 originates from point P_1.

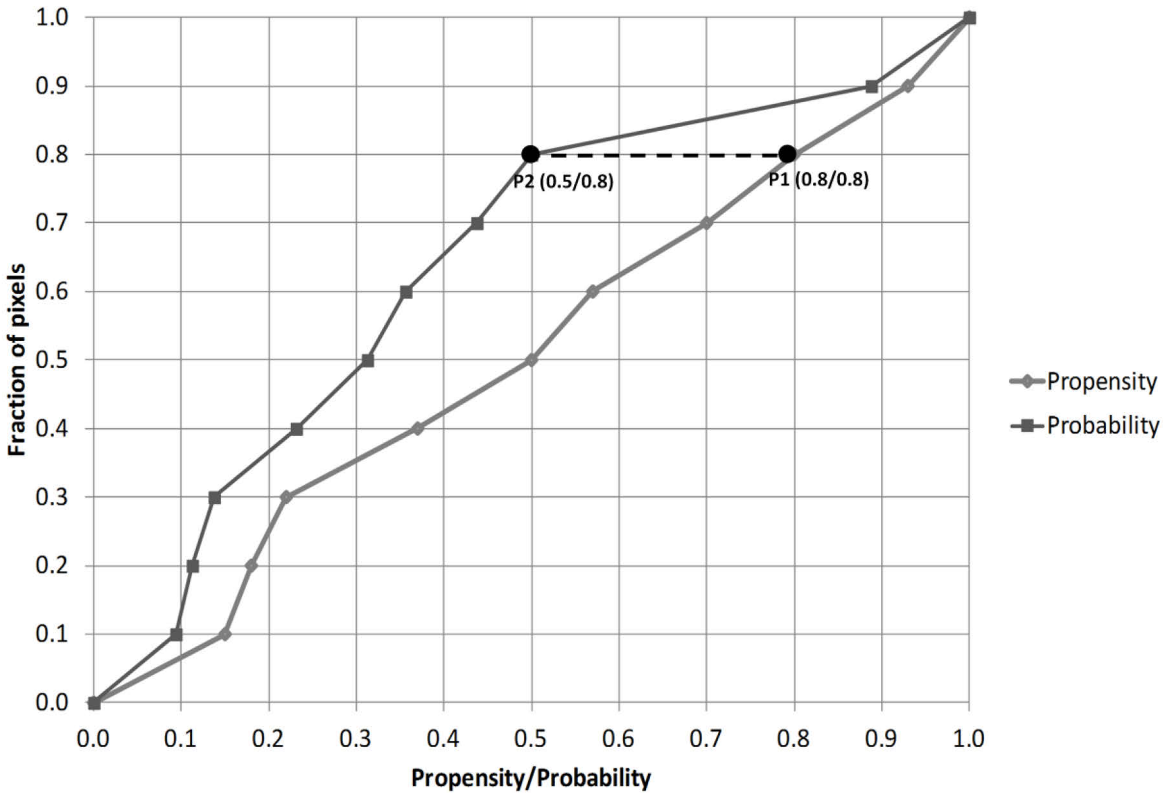


Figure III-2: Transferring propensities to probabilities (The gray and black graphs represent the propensity and probability dependent on the defined quantity of change pixels up to this value)

2.4 Quantifying Uncertainty

We develop our uncertainty measures for the probability surface following the disagreement approach of Pontius and Millones (2011) who divided the total disagreement (*TD*) of a categorical map, such as a land change map, into quantity disagreement (*QD*) and allocation disagreement (*AD*). The total disagreement is the sum of quantity and allocation disagreement. The equations are:

III-2 $$TD = M + FA = QD + AD$$

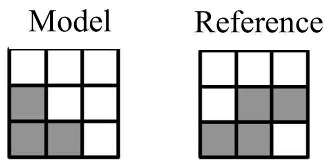
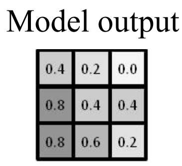
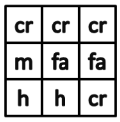
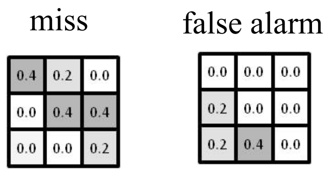
III-3 $$QD = abs(M - FA)$$

III-4 $$AD = 2 \cdot \min(M; FA)$$

where *TD* = total disagreement, *QD* = quantity disagreement, *AD* = allocation disagreement, *FA* = proportion of false alarms (“change” in the prediction map and “persistence” in the reference map) within the total number of modeled pixels, and M = proportion of misses ("persistence" in the prediction map and "change" in the reference map) within the whole number of modeled pixels.

A crisp model outcome is compared to reference data. Every difference between these two maps is either a miss ("persistence" in the model outcome and "change" in the reference data) or a false alarm ("change" in the model outcome and "persistence" in the reference data). These differences are used to calculate the disagreement measures. Table III-1 gives a simple example of how to calculate the quantity, allocation and total disagreement for two 3 x 3 pixel maps.

Table III-1: Disagreement vs. uncertainty

	Quantifying the disagreement (Pontius and Millones 2011)	Quantifying the uncertainty (this paper)
Definition	The true value is known. The disagreement is the difference to this value, i.e. it is 0 or 1 in a binary case.	The true value is not known. The uncertainty is reflected by probabilities which represent how likely it is that a given value is the true value.
Required input data	Two classified maps:  gray- “land change“, white- “land persistence“	One probability map: 
Basic measures	Misses and false alarms 	Probability to be a 
Final measures	Quantity-, allocation- and total disagreement $QD = abs(M - FA) = abs\left(\frac{1}{9} - \frac{2}{9}\right) = \frac{1}{9}$ $AD = 2 \cdot min(M; FA) = 2 \cdot min\left(\frac{1}{9}; \frac{2}{9}\right) = \frac{2}{9}$ $TD = QD + AD = \frac{1}{9} + \frac{2}{9} = \frac{3}{9}$	Quantity-, allocation- and total uncertainty $QU = 2 \cdot abs(PM - PF) = 0.18$ $AU = 4 \cdot min(PM; PF) = 0.36$ $TU = QU + AU = 0.54$
Interpretation	The model has a high/ low predicting ability.	The model is certain/ uncertain in its decision about occurring land change.

m- miss, fa- false alarm, h- hit, cr- correct rejection

QD- quantity disagreement, AD- allocation disagreement, TD- total disagreement, M- fractions of misses and the total number of modeled pixels, FA- Fraction of false alarms and the total number of modeled pixels

QU- quantity uncertainty, AU- allocation uncertainty, TU- total uncertainty, PM- probability misses: sum of probabilities of all cells to be a miss, divided by the total number of modeled pixels, PF- probability false alarms: sum of probabilities of all cells to be a false alarm, divided by the total number of modeled pixels

To calculate measures of uncertainty solely on the probability values without using reference data we need to transfer the understanding of misses and false alarms, which are the basis of the presented disagreement measures. We introduce the measure of a probability miss, which summarizes information given by the probability surface. A cell with a probability equal or greater 0.5 would be assigned to the “change” class. Therefore, the probability that the cell is a miss is 0. A cell with a probability of 0 has the highest certainty to be a “persistence” cell and the probability to be a miss is 0. The probability to be a miss increases from a given change-probability of 0 to 0.5. The slope is linear because the original probabilities are directly mirrored into the probability miss measure. This measure gives no indication about the correctness of the model; it specifies the certainty derived from the model’s outcome. A probability miss can reach values between 0 and 0.5 (Figure III-3).

The probability of a false alarm can be defined analogously. The probability of a cell being a false alarm is 0 given a probability of land change below 0.5, because this cell would be classified as a “persistence” cell in a crisp map (Figure III-4). The probability of being a false alarm is highest at 0.5 and decreases to 0 when the change-probability is 1. A probability false alarm can reach values between 0 and 0.5.

Probability misses and probability false alarms are not complementary. It is possible that a specific area has a high average value of probability misses and false alarms at the same time. However, the opposite is also possible when the model produces a high fraction of probabilities near 0 and 1.

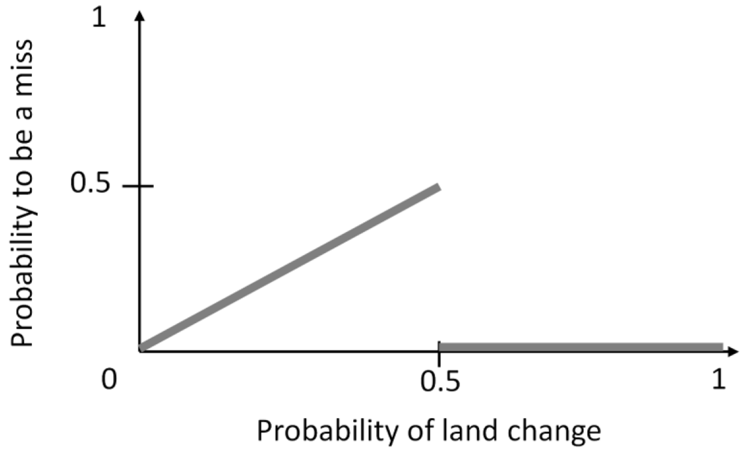


Figure III-3: Probability to be a miss dependent on the probability of land change

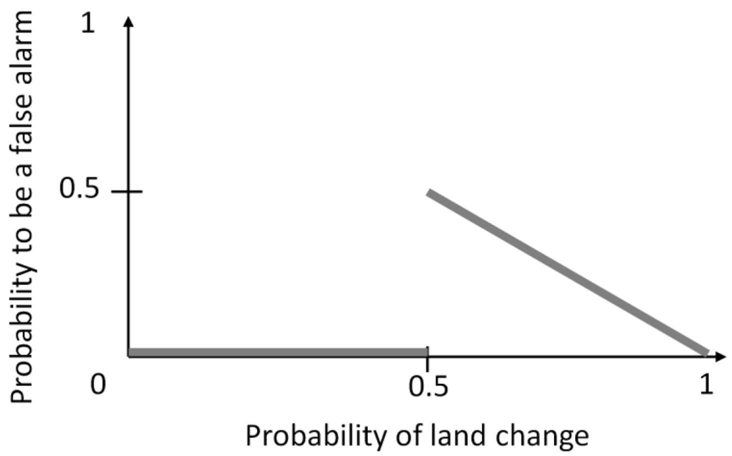


Figure III-4: Probability to be a false alarm dependent on the probability of land change

Given the definitions of a probability miss and a probability false alarm, we can define the measures of total uncertainty, quantity uncertainty, and allocation uncertainty (Equation V to VII).

III-5 $$TU = 2 \cdot (PM + PF) = QU + AU$$

III-6 $$QU = 2 \cdot abs(PM - PF)$$

III-7 $$AU = 4 \cdot \min(PM; PF)$$

where TU = total uncertainty; QU = quantity uncertainty; AU = allocation uncertainty; PF = probability false alarms-sum of probabilities of all cells to be a false alarm (following Figure III-3), divided by the total number of modeled pixels; and PM = probability misses-sum of probabilities of all cells to be a miss (following Figure III-4), divided by the total number of modeled pixels.

PM and PF are calculated by summing up all probability misses and probability false alarms (following Figure III-3 and Figure III-4) and dividing them through the total number of modeled pixels. PM and PF can reach values between 0 and 0.5. In contrast, M and FA in the disagreement equations have possible values between 0 and 1. Therefore, the uncertainty equations are multiplied with a factor twice as high as in the disagreement equations to have the same range of values. All three measures can reach values between 0 if no uncertainty exists, and 1 if uncertainty is maximal. Figure III-5 gives some examples of land change probabilities and the calculated uncertainty. The following paragraphs explain why the different types of uncertainty are as high as they are in Figure III-5.

Total uncertainty: The more pixels that are close to a probability of 0.5, the higher the total uncertainty. The total uncertainty reaches its maximum at 1 (Figure III-5, Example a). In contrast, a high fraction of values close to either 1 or 0 leads to a low total uncertainty close to 0 (Figure III-5, Example b).

Quantity uncertainty: Quantity uncertainty is dependent on the sums of probability misses and probability false alarms. A high difference between these sums leads to a high quantity uncertainty (Figure III-5, Example a). In this case most of the uncertain pixels are assigned to one land change class. If all pixels have a probability of 0.5, every pixel will be classified into the “change” category. Now assume that a difference of +/- 0.01 of the probabilities is due to randomness and has no significant meaning. We could increase the probability of half of the randomly selected pixels by 0.01 and could decrease the other half by 0.01. Based on our assumption, there would be no significant difference of the probabilities between the original probability map and the altered probability map. However, the consequences for the classification into “change” and “persistence” pixels would be substantial. In Example a, half of the pixels would be classified into “persistence” instead of “change”. This has a strong influence on the quantity of modeled land change, which is why quantity uncertainty is high.

Allocation uncertainty: A map with similar amounts of pixels equal or close to the threshold of 0.5 (both above and below) is dominated by allocation uncertainty (Figure III-5, Example d). An equal amount of these uncertain pixels is assigned to the classes “change” and “persistence”. Assuming again that a difference of +/- 0.01 of the probabilities is due to randomness, then in one possible example, one randomly chosen half of the pixels could be increased and the other half of the pixels could be decreased by 0.01 without a significant change of the probabilities. This change leads to approximately the same amount of reclassified pixels into the “persistence” class instead of the “change” class and vice versa. The quantity of land change remains approximately the same. In contrast, about half of the former “change” pixels are at a different location. Hence allocation uncertainty is high. Example c is a mixture of Example a and Example b. It has a high total uncertainty, with similar parts of quantity uncertainty and allocation uncertainty.

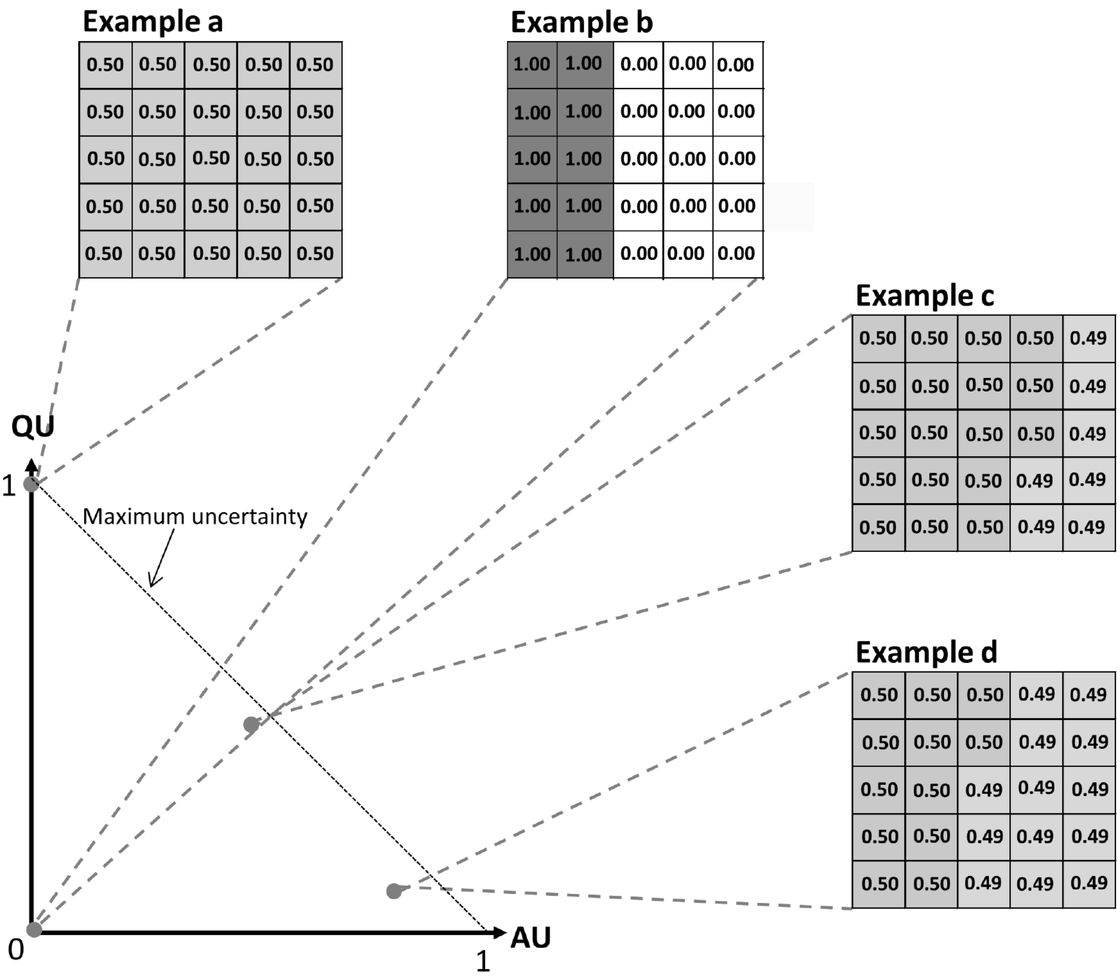


Figure III-5: Example probability maps and the respective uncertainty, QU-quantity uncertainty, AU-allocation uncertainty

Finally, we test for an empirical relationship between the existing disagreement measures (Pontius and Millones 2011) and the newly developed uncertainty measures in our case study. An existing positive relationship between disagreement and uncertainty could allow conclusions about the reliability of a model in future time steps when no reference data exist and only the calculation of uncertainty is possible. We calculate the measures for 51 municipalities in the study area and compare the quantified uncertainty with the disagreement using an ordinary least squares regression model (Equation VIII). The quantity of land change is obtained from the real change data for every municipality. The spatial allocation of the probabilities is done for every pixel. Subsequently, the disagreement and uncertainty measures are calculated and aggregated on municipality level.

III-8 $$U_i = a + b \cdot D_i + e_i$$

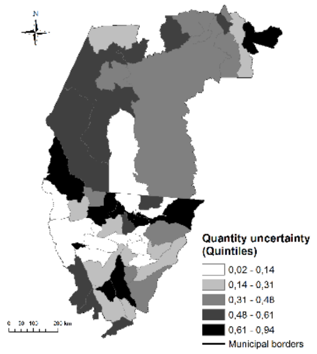
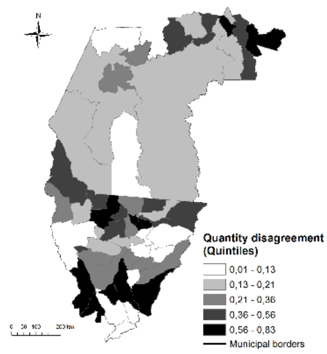
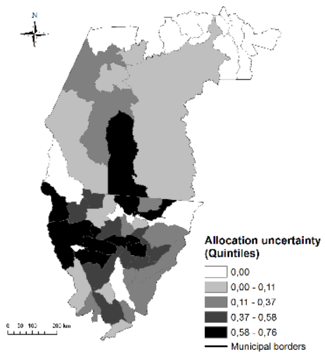
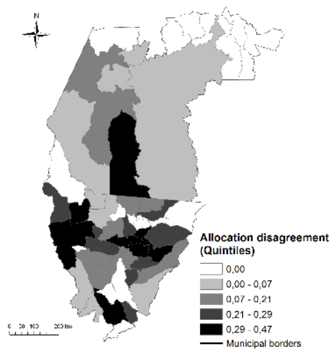
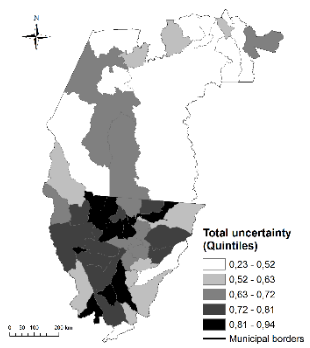
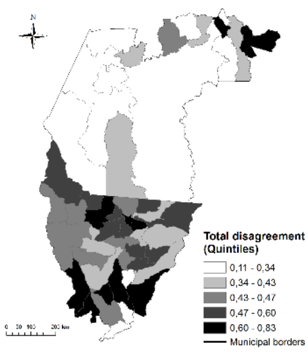
where U_i = Uncertainty (quantity, allocation or total) in municipality i, D_i = Disagreement (quantity, allocation or total) in municipality i, a, b = Estimated regression coefficients, and e_i = Residue

3 Results

From the probability maps of deforestation for our study area we derived maps of probability misses and probability false alarms Figure III-6 and Figure III-7. The pixel-based probability values were aggregated for every municipality. A high value in Figure III-6 means that the average probability of a pixel to be miss is high. In contrast, high values in Figure III-7 highlight those municipalities where the average probability of a pixel to be a false alarm is high. Based on the probability misses and probability false alarms, the uncertainty measures were calculated for every municipality (Table III-2). The total uncertainty was differentiated into quantity uncertainty and allocation uncertainty. We find a high quantity uncertainty especially in the western parts of the state Pará and the northern municipalities of the state Mato Grosso. Quantity uncertainty values of above 0.8 are concentrated in these areas of Mato Grosso. A dominant part of the uncertainty is due to the probability misses (Figure III-4) which exceed the sum of probability false alarms. That means, the chance of omitting real change in the modeling outcome is high. The areas with a low quantity uncertainty are allocated nearby the described municipalities with high values of quantity uncertainty. Municipalities in the center of Pará and Mato Grosso reach the lowest values of quantity uncertainty. In contrast, most of these areas have a high allocation uncertainty, up to 0.76. The majority of municipalities with a low allocation uncertainty are situated at the edges of our study area, especially in the northern parts.

The disagreement measures which were additionally calculated for the investigated time period are shown in Table III-2. The spatial distributions of the disagreement measures show patterns similar to the patterns of the uncertainty measures. The relationship between disagreement and uncertainty is also emphasized by the positive correlations of 0.68 to 0.85. The modeled disagreement to the reference data is higher in areas where the calculated uncertainty is high as well. This tendency is significant at the 0.001 level following the Wald statistics for each of the three uncertainty measures.

Table III-2: Disagreement and uncertainty of modeled land change in the Brazilian case study: every gray shade represents a quintile (dark gray = high disagreement/uncertainty)

Uncertainty	Disagreement	Relationship
Quantity uncertainty  	Quantity disagreement  	R: 0.7442*** R^2: 0.5538 b: 0.6436***
Allocation uncertainty  	Allocation disagreement  	R: 0.8476*** R^2: 0.7186 b: 0.4347***
Total uncertainty  	Total disagreement  	R: 0.6838*** R^2: 0.4676 b: 0.7153***

R - Correlation coefficient; R^2 - Coefficient of determination; b - empirical slope in a linear regression ($U_i \sim a + b \cdot D_i + e_i$); , *** significant at 0.001 level

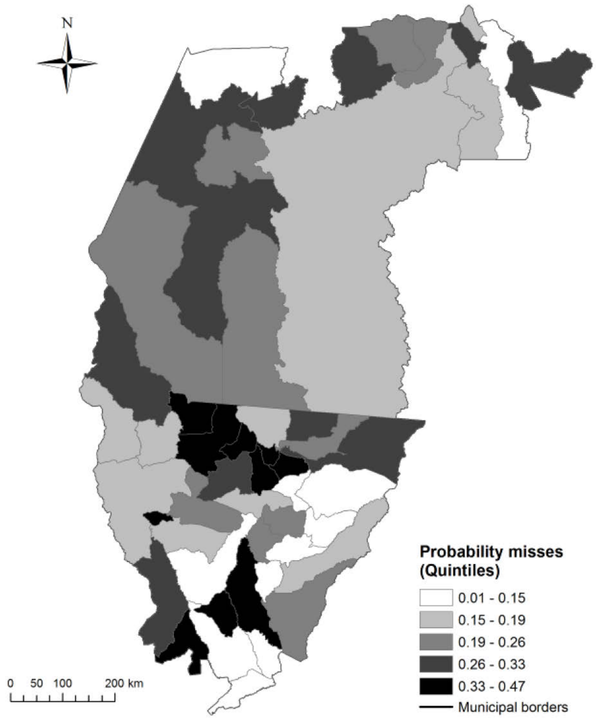


Figure III-6: Probability misses in the Brazilian case study, 2002-2005: every gray shade represents a quintile (dark gray = high values)

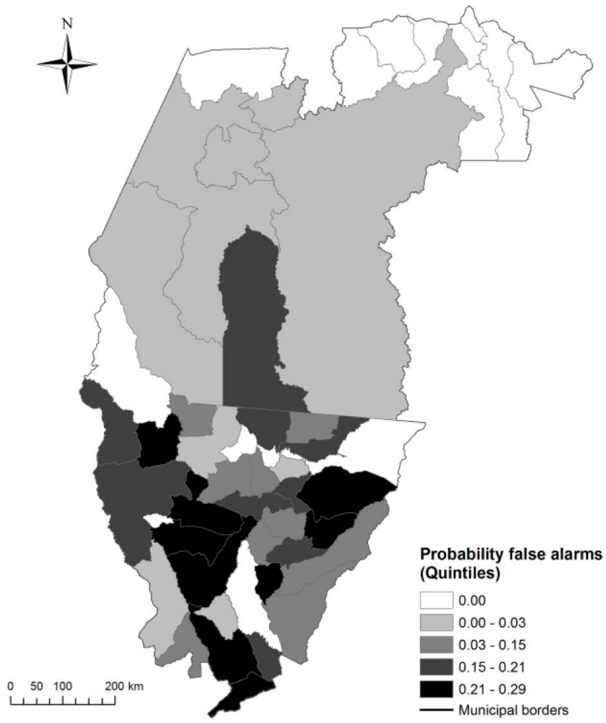


Figure III-7: Probability false alarms in the Brazilian case study, 2002-2005: every gray shade represents a quintile (dark gray = high values)

4 Discussion

Our findings show that the characterization of uncertainty for land use change modeling outcomes is possible by using the proposed measures of total uncertainty, quantity uncertainty and allocation uncertainty. They allow for the assessment of the degree of uncertainty associated with the probability-based outcome of a land change model. We test our uncertainty approach in a BBN modeling experiment to analyze deforestation. At the same time, our proposed measures are neither limited to a specific land change modeling technique nor to a specific kind of land change. Other modeling approaches such as Neural Networks or Classification and Regression Trees (Tayyebi, A. and Pijanowski 2014) which produce continuous output maps are additionally suitable. Irrespective of the chosen modeling technique, it is useful to carefully test the empirical relationship between uncertainty and disagreement in the chosen case study before applying the approach to future time steps.

4.1 Understanding uncertainty

The suggested uncertainty measures are a beneficial approach to evaluate the performance of a land change model by complementing established accuracy measures of goodness of modeling results. However, an entirely certain land change model which exclusively assigns probability values of either 0 or 1 can be a worthless model. Even with no uncertainty in the modeling results, every pixel value can be wrong. Based on the developed measures, we can only state whether a model is certain or uncertain. The decision about its goodness cannot be derived solely from its uncertainty. The relationship between the quantified uncertainty and disagreement in a known time step, however, may be used as a first hint for the disagreement in predicted time steps. In our case study, we quantified the relationship between disagreement and uncertainty. The expectation that a high disagreement occurs if the uncertainty is high is verified in this case study. Therefore, a high uncertainty in an unknown region/ time step without reference data would lead us to the conclusion that the model delivers unreliable results in this case. However, the strength of the empirical relationship for the different uncertainty and disagreement categories is dependent on the characteristics of the land change models and the chosen case study. The relationships should be tested in every case study in known time steps whenever possible before applying the uncertainty approach in unknown time steps. Therefore, reference data for the known time steps are useful to test the applicability of the uncertainty concepts in future time steps without reference data. However, a positive relationship between disagreement and uncertainty in known time step does not guarantee a similar relationship in the future. Non-stationary behavior of land change systems could influence the predicting model performance which is often based on a "business as usual scenario" (Müller *et al.* 2014).

The calculated uncertainty values give an indication whether a modeling approach is reliable for the purpose of the user. However, it is not possible to define an uncertainty value which differentiates between beneficial and useless land change models. Beside the utility of the land change model, the reached disagreement additionally depends on the complexity and randomness of the modeled process. An exclusively random process cannot be predicted, even if the land change model is reliable. A suitable way to decide if a land change model is useful is to use an appropriate reference model. One example for the application in the land change community is to use a land change model which only includes the "distance to previously changed areas" as an independent variable.

4.2 Separating uncertainty

Separating the uncertainty into quantity and spatial components is particularly useful, because several land change models operate on these two dimensions. E.g. the CLUES model (Verburg *et al.* 2002) connects the land change demand for the whole study area with the spatial allocation of land change on the pixel level. Another example is the LandSHIFT model (Schaldach *et al.* 2011) which combines the regional and country levels. After the calculation of the distinct uncertainty measures, the land change modeler is able to decide where to improve the land change model.

For our case study, we can identify areas with high quantity uncertainty. In combination with the illustration of probability misses and probability false alarms, it is possible to infer whether the model is prone to underestimate or overestimate real land change. This is most likely due to the effect of global variables or other variables which are important to quantify the demand of land change for the whole investigated region. Within our modeling exercise, the socioeconomic variables are constant within a municipality. Therefore, they influence solely the amount of deforested pixels assigned to a specific municipality. Model improvements which aim to reduce uncertainty in the model outcome should address these variables, e.g. by including alternative data sources or by disaggregating coarse scale variables to a finer resolution (Krüger and Lakes 2014).

Allocation uncertainty helps to identify where the model is uncertain about the placement of land change, even if the model input includes the exact amount of real land change. Variables which vary within a certain area mainly influence this uncertainty. Model improvements should focus on these variables and their parameters in the land change model when the allocation uncertainty is dominant.

4.3 Limitations

The described approach has some limitations. The uncertainty measures are developed to address binary land change problems only; however, at the same time such change studies are widely used not only for deforestation but also urbanization analyses. Moreover, the selected linear regression analysis is only one possibility to analyze a relationship between uncertainty and disagreement. This approach is limited in terms of a possible non-linear behavior. Machine learning approaches, such as Neural Networks, are alternatives. To investigate the relationship between disagreement and uncertainty, we decided to divide the study area using administrative boundaries. Many alternatives are possible, e. g. separating the study area into subareas with homogeneous spatial characteristics.

5 Conclusions

Our study contributes to an up-to-now rarely studied field of uncertainty analysis in land change modeling and spatial modeling in general. The measures we propose are calculated for modeled probabilities and are applicable when there is no reference information available. The calculated total uncertainty is differentiated into uncertainty which is caused by the wrong quantity of modeled land change and uncertainty caused by the incorrect spatial placement of land change. We applied this approach to a deforestation case study of Brazil and demonstrated the usability of this concept along with established disagreement measures. We propose to test for a relationship of the disagreement measures and calculated uncertainty in further case studies. In case of an existing relationship, the uncertainty of land change projections is a hint for the reliability of land change models. This study shows that quantifying uncertainty in land change modeling outcomes provides an important asset for land change modelers to gain new insights for revisiting and adjusting their land change models.

Future work will focus on the possibilities to compare different projections. Two projections can be very similar even if the underlying assumptions are substantially different (Pontius and Neeti 2010). Uncertainty in the projection outcome maps may hide a high fraction of such differences.

Acknowledgments

We thank Florian Gollnow and Andreas Rabe for helpful comments and fruitful discussions that helped to improve this manuscript. The authors additionally thank Dennis Funke for data preparation work. We additionally thank the anonymous reviewers who helped us to improve the earlier versions of the article by giving a variety of beneficial remarks and suggestions.

Funding

The presented work is supported by Deutsche Forschungsgemeinschaft [GRK 1324/1, METRIK].

Chapter IV:
How similar are two land change projections?
Submitted manuscript

Carsten Krüger and Tobia Lakes

Abstract

Contrasting alternatives of spatially explicit projections from land change models are often used to reveal the range of future development paths. The underlying qualitative storylines frequently vary substantially. The derived maps, however, are not necessarily different, because uncertainties may dominate differences associated with storylines. This article presents a methodical framework to analyze how similar two spatially explicit land change projections are. We consider two outputs of the land change model: the classified land change maps and the underlying probability maps of land change. We apply and develop measures which separate the disagreement into spatial and quantity components. Additionally, we define a reference which helps to decide if two projections are similar or not. Our proposed set of measures allows to compare different land change projections in a quantitative way. Such information is of importance for decision-makers and scientific modeling chains relying on different projections of land change and beyond.

1 Introduction

A comparison between today's Earth surface and the Earth's surface 100 years ago yields very distinct pictures. Deforestation, urbanization, and agricultural expansion and intensification are some examples of processes which currently shape the landscape. The majority of the change processes are induced by humans to satisfy their needs for food, shelter, mobility, or recreation (Foley *et al.* 2005). Each of the mentioned processes has consequences, some of which have severe implications for the vitality of the land system, for example a loss of biodiversity, flooding of impervious surface or overfertilization of soils. Land change models can give some guidance about how to balance out the trade-offs between satisfying human needs on the one hand and mitigating the severe ecological implications on the other. These models estimate which factors are most important for the observed land change in the past or present, and project where land change is likely to occur in the future (Brown *et al.* 2013). Such estimations cannot represent the reality 100 %; a certain degree of uncertainty is always included. It is crucial to know and communicate the degree of uncertainty for political and economic decision-makers.

Typically, a land change model is calibrated with land cover or land use maps representing at least two consecutive time steps t_1 and t_2. Subsequently, the calibrated model is used to project future land change up to the time step t_3. Frequently, it is assumed that locations which were suitable for land change in the past are suitable for land change in the future as well. The result of the calibration process is a continuous, area-wide output raster where a specific value is assigned to each pixel (Müller *et al.* 2012; Mas *et al.* 2014). Depending on the modeling strategy, the research discipline, and the individual researcher, this value is interpreted as suitability, propensity, transition potential or probability. This continuous raster is then used for the spatial allocation of future land change pixels (Gollnow and Lakes 2014; Krüger and Lakes 2014). The spatial allocation of a binary land change case means that one part of the pixels will be classified into the "land change" category for the future time step t_3 and the other part will be classified into the "land persistence" category. Frequently, this classification is performed by using a given number of pixels that have to change, the so called "land change demand" (Verburg *et al.* 2002; Schaldach *et al.* 2011). Such a demand can be derived from different sources, e.g. economic models about market prices or based on different assumptions. The demand is related to the quantity of land change. Since the spatial allocation of land change pixels and the land change demand are frequently the subject of separated modeling components (Asselen and Verburg 2013), it is reasonable to evaluate the uncertainty of these two components separately as well (Krüger and Lakes forthcoming).

This article addresses the evaluation of land change projections, i.e. the specification of which pixels will be land change pixels in future time steps. The related output of the land change model cannot be compared with true land change data to assign the accuracy afterwards, because the future real land change does not exist. An alternative to evaluate the model output is to compare two different projections (Pontius and Neeti 2010; Alcamo *et al.* 2011). In both projections we assume reasonable storylines, model settings and underlying assumptions. In this case, we receive two raster datasets as model outputs which project the location and quantity of land change for a future time step t_3. The question whether one of these projections is correct or reaches a reasonable accuracy cannot be answered until t_3 becomes reality. However, we can obtain valuable information by comparing the two projections. If the projections are similar, we will increase our certainty about what will happen in the future. In contrast, if we get two substantially different projections, our uncertainty about the future is higher. Opposing different projections serves as an uncertainty range (Gutierrez-Velez and Pontius 2012). This leads to the questions of how to compare the two projections and which criteria can be used to decide if the projections are similar or not.

In previous work, comparisons of different projections were frequently completed without giving any quantitative information about their differences (e.g. Verburg *et al.* 2010). Another established approach is to compare the classified output maps (e.g. Sohl *et al.* 2012). In this approach, the continuous model output is classified into a binary set of "land change" and "land persistence" pixels. Typically a confusion matrix is calculated by assessing the four possible states: a) land change in both projections P_1 and P_2, b) land persistence in both projections, c) land change in P_1 and land persistence in P_2, and d) land persistence in P_1 and land change in P_2. From this confusion matrix one can calculate the total disagreement, quantity disagreement, and allocation disagreement (Pontius and Millones 2011). This approach allows the differentiation whether the majority of the total disagreement is due to a different amount of land change pixels in both projections, or due to a different spatial placement.

Figure IV-1 a) and b) give a simple example for comparing two projections following the disagreement approach described above. This figure provides the output of two projections. Both 5 x 5 pixel maps are classified into the two classes: "land change" in black and "land persistence" in white. P_2 has two more land change pixels than P_1 and thus the quantity disagreement is 2. Additionally, every land pixel which is predicted to be "land change" in P_1 is also "land change" in P_2. There is no spatial shift between the two projections. This means that the allocation disagreement is 0. The total disagreement is the sum of both described components, which, in the example, is 2. Expressed as a fraction of the whole number of pixels, it is 0.08. The two maps seem to be quite similar.

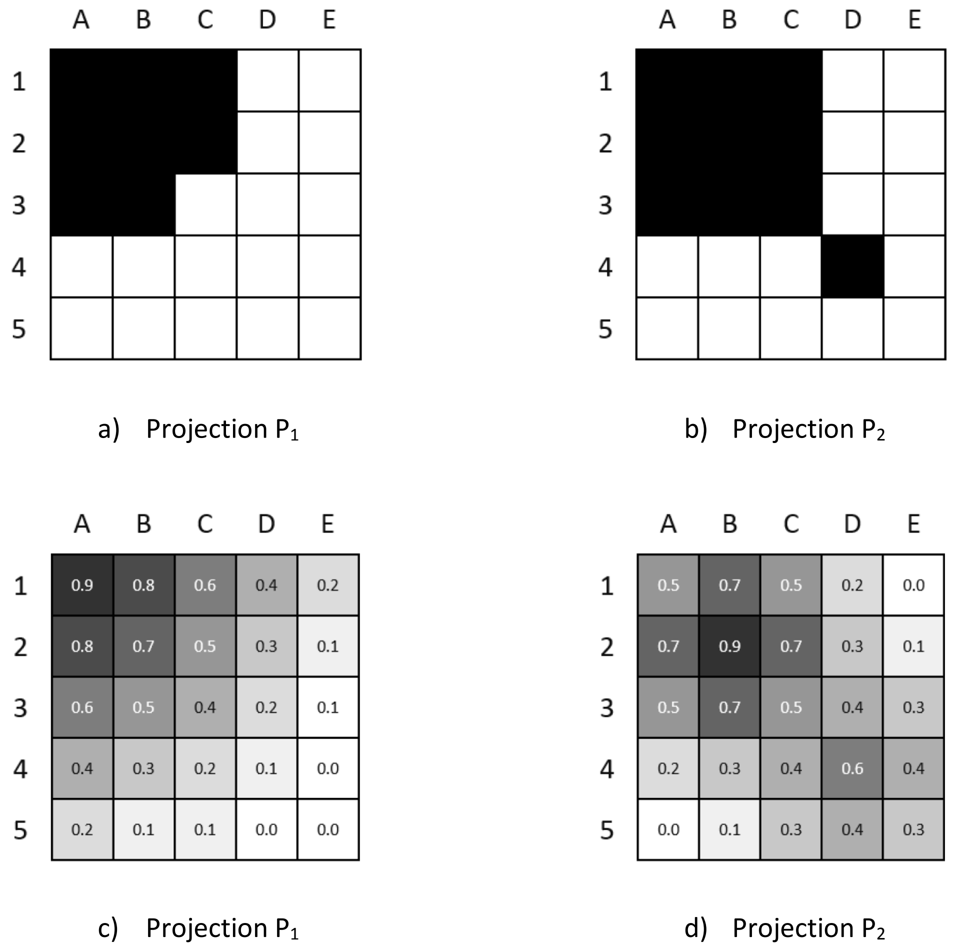


Figure IV-1: a and b) example comparison of two land change projections: binary, c and d) continuous probability values

The original model outputs before classifying the pixels into two classes were continuous raster maps with pixel values between 0 and 1. We assume that these values represent probabilities of future land change, with 1 as a modeled probability of 100 %. The classified maps are obtained from the probability maps by assigning the "land change" category to every pixel with a probability equal or above 0.5, and by assigning the remaining pixels to the "land persistence" category. Figure IV-1 c) and d) give the original continuous output maps of the example introduced above. When looking at these maps, different patterns seem to be obvious in contrast to the comparison of the classified maps. There is no clear answer to the straightforward question if the two projections are similar. For example, pixel C_3 is different in the classified maps. At the same time, it has a relatively small difference in terms of probability of 0.1. In contrast, pixel A_1 is classified in the "land change" category in both maps. However, the probabilities differ by 0.4.

The objective of this article is hence to analyze the question of similarity of different land change projections. Therefore, the first aim is to develop and apply a set of measures that covers quantity, spatial, and total disagreement of both the classified maps and the continuous maps. The second aim is to analyze if there is certain disagreement threshold which helps to distinguish between similar and different projections. In the remaining chapters of the article we first describe the respective measures and then apply the approach to a land change modeling case study using SimWeight (Sangermano *et al.* 2010) which is implemented in the Land Change Modeler (IDRISI, Clark Labs) to project deforestation in Brazil.

2 Methods

To analyze the similarity of land change projection maps we first describe the measures which we use and develop. We than explain the case study for which we test the approach.

2.1 Measuring the disagreement

The confusion matrix which is the basis for a variety of agreement and disagreement measures (e.g. Hagen-Zanker 2009, Mas *et al.* 2013; Pontius *et al.* 2013) relies on the comparison of pixels. These pixels are at the same position in two classified maps. In a binary case, a pixel is either coded as 1 which represents "land change" or as 0 which represents "land persistence". The comparison of two pixels leads to 4 possible combinations which are summed up in the confusion matrix for the whole number of compared pixels (Figure IV-2 a). The diagonal from the upper left to the lower right represents the agreement, whereas the diagonal from the lower left to the upper right represents the disagreement. The two disagreement fields in the binary case can be used to calculate the disagreement differentiated into quantity and spatial components following Pontius and Millones (2011).

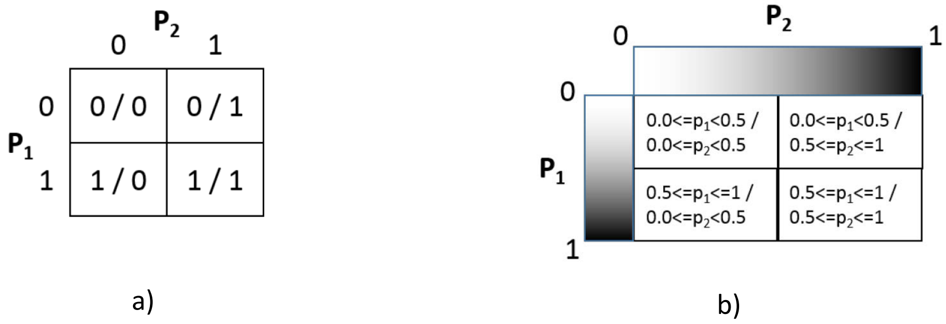


Figure IV-2: Binary confusion matrix: a) classification of the binary maps, b) implicit classification of the continuous model outputs

We assume probabilities as the original model output. These probabilities (*p*) have 0.5 as the threshold to decide which category will be assigned to a pixel in the classified map. To derive the described confusion matrix in Figure IV-2 a, the matrix in Figure IV-2 b is used implicitly. Therein, original substantial differences between two pixels can be condensed as an agreement when e.g. $p_1 = 0$ and $p_2 = 0.49$. In comparison, slight differences between two pixels can be emphasized as a difference when e.g. $p_1 = 0.49$ and $p_2 = 0.50$.

The idea behind the measures based on the continuous output is to use all the information given by the probabilities as it symbolized in Figure IV-3 a). Two pixels with probabilities near the diagonal from the upper left to the lower right are similar, whereas disagreement increases with increasing distance from this diagonal. We get a continuous disagreement space when using the continuous probabilities as basis for the comparison.

As described above, the comparison of two classified pixels delivers four possible outputs, of which only one occurs. Instead, the comparison of two pixels with probabilities leads to four different probabilities as outputs for pixel i. We have two independent probability distributions of the compared projections P_1 and P_2. A probability of p_{1i} that land change will occur in P_1 implies that no land change will occur in P_1 with a probability of $1 - p_{1i}$. Simultaneously, the probabilities for P_2 are p_{2i} and $1 – p_{2i}$. Multiplying the probabilities leads to the four following joined probabilities:

1) $p_{1i} \cdot p_{2i}$: The joined probability for future land change in both projections
2) $p_{1i} \cdot (1 - p_{2i})$: The joined probability for future land change in P_1 and land persistence in P_2

3) $(1 - p_{1i}) \cdot p_{2i}$: The joined probability for future land persistence in P_1 and land change in P_2

4) $(1 - p_{1i}) \cdot (1 - p_{2i})$: The joined probability for future land persistence in both projections

These probabilities can be assigned to a matrix similar to the confusion matrix which we name "probability confusion matrix" (Figure IV-3 b). Again, the diagonal from the lower left to the upper right represents the disagreement. The two fields in this diagonal are used to calculate the probability disagreement measures as described in detail below.

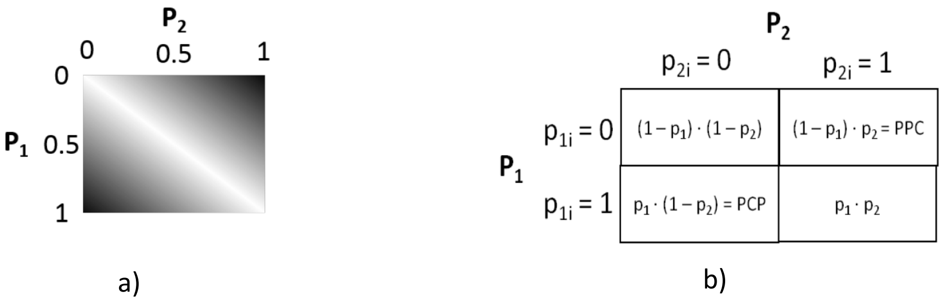


Figure IV-3: Probability confusion matrix: a) continuous space, b) classified

We use the following notation for the mathematical explanations given below:

N	Number of pixels
TD	Total disagreement
QD	Quantity disagreement
AD	Allocation disagreement
TPD	Total probability disagreement
QPD	Quantity probability disagreement
APD	Allocation probability disagreement
CP	Sum of pixels with predicted "land change" in projection P_1 and predicted "land persistence" in projections P_2

PC	Sum of pixels with predicted "land persistence" in projection P_1 and predicted "land change" in projections P_2
p_{ki}	Probability for future "land change" in pixel i following projection k (1, 2)
$(1 - p_{ki})$	Probability for future "land persistence" in pixel i following projection k (1, 2)
PCP	Probability of change in P_1 vs. persistence in P_2: Average of $p_{1i} \cdot (1 - p_{2i})$ over all pixels
PPC	Probability of persistence in P_1 vs. change in P_2: Average of $(1 - p_{1i}) \cdot p_{2i}$ over all pixels

The existing disagreement measures for classified maps (Chen and Pontius 2010) can be calculated as follows:

IV-1 $$QD = abs(CP - PC)$$

IV-2 $$AD = 2 \cdot min(CP; PC)$$

IV-3 $$TD = QD + AD$$

The developed probability disagreement measures refer to the probability confusion matrix instead of the original confusion matrix. Therefore, we replace *CP* and *PC* and receive the following equations:

IV-4 $$QPD = abs(PCP - PPC)$$

IV-5 $$APD = 2 \cdot min(PCP; PPC)$$

IV-6 $$TPD = QPD + APD$$

QPD, *APD* and *TPD* can be rewritten as:

IV-7
$$QPD = abs(\sum_i p_{1i} \cdot (N - \sum_i p_{2i}) - (\sum_i p_{2i} \cdot (N - \sum_i p_{1i}))) = abs(N \sum_i p_{1i} - N \sum_i p_{2i})$$

IV-8
$$APD = 2 \cdot \min(\sum_i p_{1i} \cdot (N - \sum_i p_{2i}); \sum_i p_{2i} \cdot (N - \sum_i p_{1i}))$$

All three probability disagreement measures are scaled between 0 when the probability disagreement is lowest and 1 when the disagreement is maximal. Figure IV-4 depicts 5 examples showing a comparison between two 3 x 3 pixel probability maps. The further the distance to the origin of ordinates is, the higher the total probability disagreement. The maximum of 1 is given by a dotted diagonal. The total probability disagreement approaches the maximum when the difference in probabilities for each pixel is close to one. That means that the majority of pixels has a probability close to one in one prediction and a probability close to 0 in the other prediction (Example 1 and 5).

The total probability disagreement is separated in its components quantity probability disagreement and allocation probability disagreement. Quantity probability disagreement is predominant when the probabilities of “land change” are substantially higher in one projection in comparison to the second projection (Example 1 and 2). In comparison, allocation probability disagreement is substantial when most of the pixels with the highest probabilities are at different locations in both projections (Example 3, 4 and 5). Especially example 5 has a high allocation probability disagreement close to 1, because pixels with a probability of land change of 1 are at completely different locations in both projections.

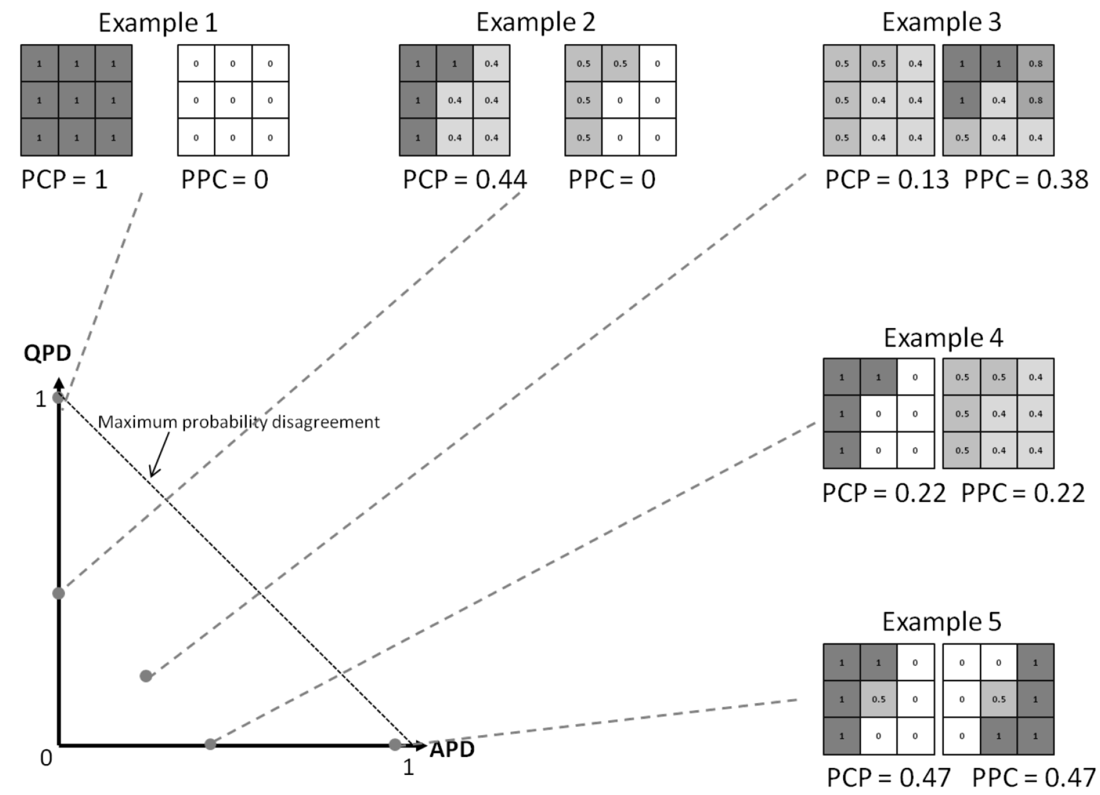


Figure IV-4: Disagreement space of the probability disagreement measures and five examples

2.2 Visualizing disagreement

We summarize the 6 explained and developed measures of disagreement in a spider chart (Figure IV-5). One graph represents the calculated measures of disagreement for two different projections. Values close to the center of the chart represent sparse disagreement, whereas disagreement increases with increasing distance from the center. The blue graph depicts the example from chapter 1 in this article. The graph highlights that the disagreement of the classified maps is low; however, the disagreement of the probability maps is high. The allocation probability disagreement is considerable, whereas the allocation disagreement is 0.

Disagreement values for the proposed measures vary according to the fractions of the class sizes. The disagreement will be relatively low if one of the two land change classes is small. For example, if both projections assume a land change of 0.04, then the allocation disagreement and the total disagreement have their maximum at 0.08. For that reason, an appropriate reference comparison is needed to adequately represent the differences of the two projections. We choose a comparison with a random raster as a reference comparison. A random raster in this case means a random allocation of a land change demand which was observed in the past. We divide each resulting disagreement value by the value resulting from the comparison with this random raster. Therefore, every ordinate in the spider diagram is scaled to the random reference in terms of a fraction of disagreement to the disagreement with a random raster. A disagreement of 1 in the spider chart means that P_1 has the same disagreement with P_2 as with a random raster. In that case, the disagreement between P_1 and P_2 is substantial. The red graph represents the former blue graph adjusted to the described ordinate division. In the following paragraphs, we refer to adjusted differences when we write about differences of two maps in general.

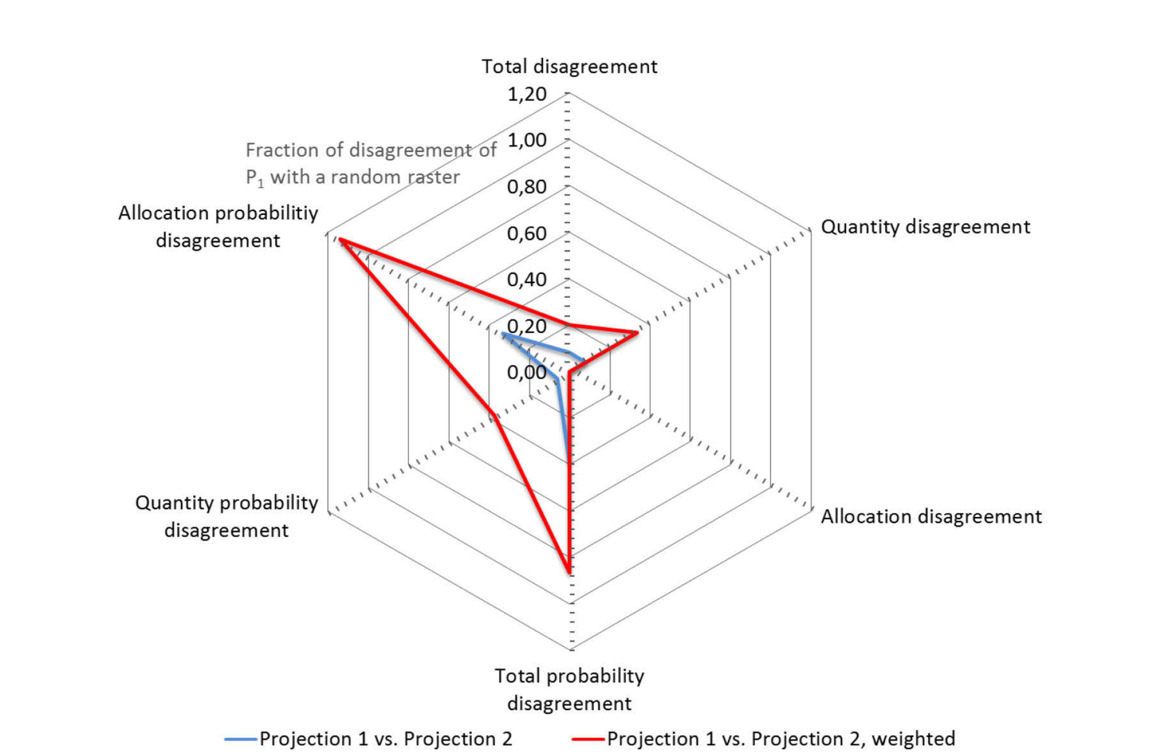


Figure IV-5: Example disagreement spider chart

2.3 Study area, data and land change model

We chose a deforestation case study region in the Brazilian Amazon to test our approach for comparing land change projections. The area is situated in the Brazilian Amazon, in the two states Pará and Mato Grosso. During the last 15 years, the whole Brazilian Rainforest had yearly deforestation rates of between 4,656 km^2 in 2012 and 27,772 km^2 in 2004 (Börner *et al.* 2015), which is nearly the area of Belgium. The loss of original forested area has had severe ecological impacts, such as overfertilization, soil erosion, loss of the biodiversity, and increasing atmospheric CO_2 level (Fearnside 2008; de Espindola et al. 2012).

We use PRODES land cover data (Instituto Nacional de Pesquisas Espaciais (INPE) 2013) from 2002 and 2005 to calibrate our land change models. The predictions are made for the year 2011. We decided to keep the land change models as simple as possible, since the major aim of this article is to present an approach to compare different land change projections. For that reason we use three different land change models, each one of which uses one of the following explanatory variables:

a) “Fraction of large farms”: Fraction of the number of farms with a size > 1000 ha to the whole number of farms

b) “Distance to populated places”: Euclidean distance to villages/cities

c) “Elevation”

These variables are differently shaped and lead to distinct projected patterns of land change. The data processing and the sources of these data is described in Krüger and Lakes (2014).

We select the SimWeight (Sangermano *et al.* 2010) algorithm which is implemented in IDRISI. SimWeight is an algorithm similar to the k-nearest neighbor. It estimates a continuous transition potential for every unknown pixel dependent on the values of the known pixels. Therefore, the weight of a known pixel increases with decreasing distance to the unknown pixel. After the calibration of the land change model, we receive transition potentials as a model output which reflects where future land change is more probable. A pixel with a higher transition potential has a higher probability of future land change as well. However, it is not possible to conclude that a transition potential of 0.4 means a probability of future land change twice as high as a transition potential of 0.2. Without specifying the expected amount of future land change, these transition potentials represent only where future land change is more probable, but not how probable. Knowing or estimating the amount of future land change allows the transfer of the transition potentials to probabilities (Krüger and Lakes forthcoming). The amount of future land change is in this case obtained from real change data for the time period of 2006-2011.

2.4 Software Implementation

The necessary raster calculations are implemented in an ArcGIS 10.2 (Esri) model (Figure IV-6). The model uses transition potentials of two projections as input data. Additionally, one threshold for each of the projections is required. The threshold specifies the transition potential value which separates the "land persistence" and "land change" pixels. The threshold is used to calculate probabilities with a threshold of 0.5 (Krüger and Lakes forthcoming). In a next step, the two probability maps are used to calculate the joined probabilities $p_{1i} \cdot (1 - p_{2i})$ and $p_{2i} \cdot (1 - p_{1i})$ for every pixel of the study area. These probabilities are summed up to PCP and PPC and subsequently used to calculate the uncertainty measures.

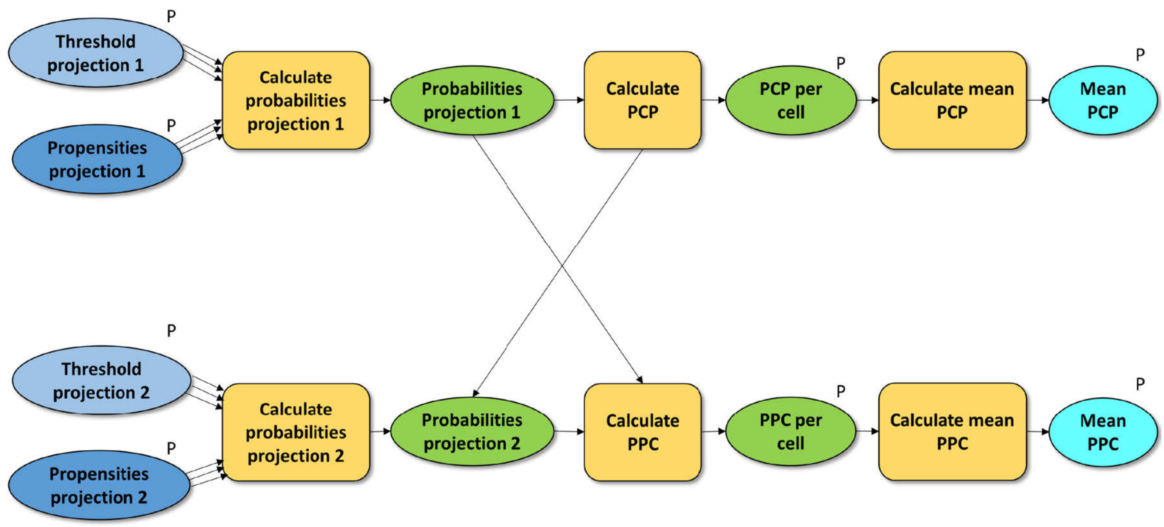


Figure IV-6: ArcGIS Modeling framework

As described above, the absolute values of the disagreement and uncertainty measures are dependent on the amount of "land change". Therefore, we relate the comparison of two projections to a comparison with a random raster. The values are summed up in the spider chart.

3 Results

After applying the SimWeight algorithm to our different land change models, we received maps of transition potentials (Figure IV-7 a and b). The transition potential maps are converted into probability maps and in maps with the joined probabilities of $p_{1i} \cdot (1 - p_{2}i)$ and $p_{2i} \cdot (1 - p_{1i})$ afterwards. Figure IV-8 a and b give these probabilities for p_1 = "fraction of large farms" and p_2 = "distance to populated places". These maps allow the investigation of where the difference between the given probabilities is largest (red color).

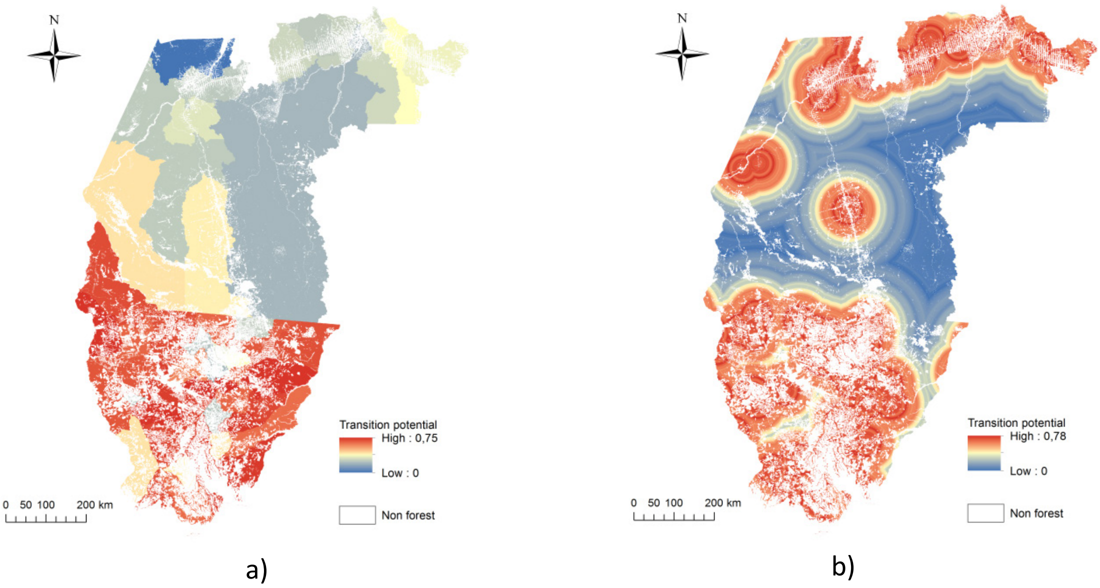


Figure IV-7: Transition potential for the land change models based on a) "fraction of large farms" and b) "distance to populated places"

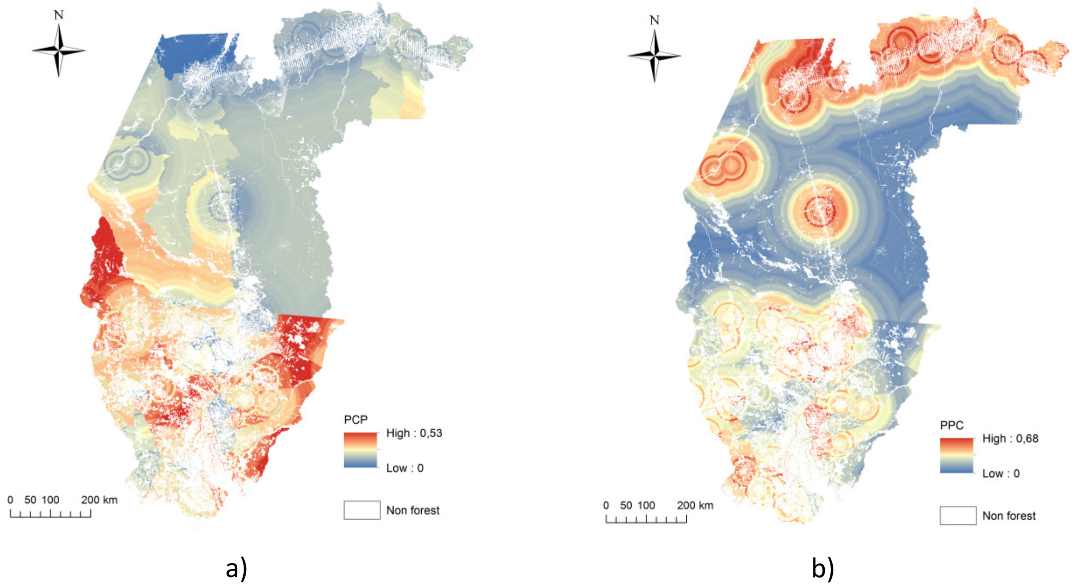


Figure IV-8: a) PCP and b) PPC of the projections based on "fraction of large farms" and "distance to populated places"

The probability maps and additionally constructed binary land change projections are used to calculate the measures presented in this article. The measures are summarized in a spider chart (Figure IV-9 and Figure IV-10). In Figure IV-9 the projection based on the variable “fraction of large farms” is compared with the projections based on “distance to populated places” (C_1) and “elevation” (C_2). The comparison with a random prediction is the reference comparison (C_R). C_R is highlighted by a black line and has a value of 1 in every dimension. A value of 0.5 in the spider chart means that one projection has half of the disagreement with another projection in comparison to the disagreement with a random raster.

The red graph shows the comparison C_1. The graph is close to the comparison C_R with a random projection in four disagreement dimensions close to the comparison C_R with a random projection. That means that there are sharp differences between these two projections. The quantity disagreement is the same in C_1 and C_R because all projections rely on the same specified land change demand. (The demand is the same in all projections. Therefore, the number of land change pixels is the same. A quantity disagreement in comparison C_R of 0 leads to a division by 0 and to a non-defined value for the weighted quantity disagreement. We set the value to 0 because the unweighted quantity disagreement is 0 in every comparison.) The quantity probability disagreement is the only dimension with weaker differences in C_1. The blue graph represents comparison C_2. The differences in C_2 are weaker in terms of allocation disagreement and total disagreement than the difference in

C_1 described before. The difference with the random comparison is higher. At the same time, the difference in C_2 in terms of the probability disagreement measures is higher than C_1. The binary and continuous disagreement measures provide different results.

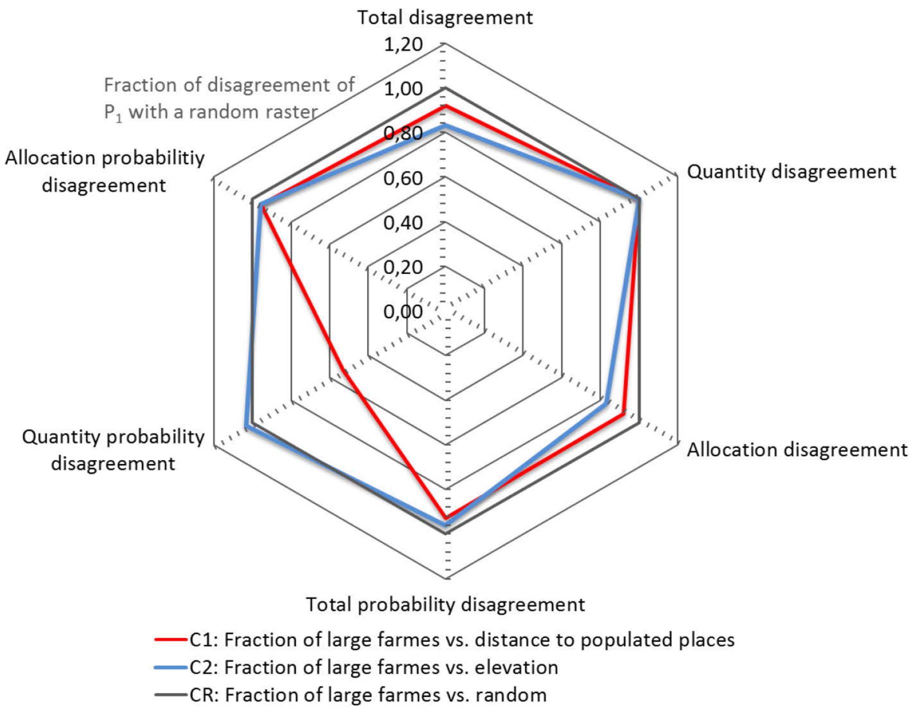


Figure IV-9: Comparison of land change projections with different explanatory variables

Figure IV-10 shows the comparison of a projection based on the variable “fraction of large farms” with a projection based on the same explanatory variable; however, the specified land change demand is triple as high as in the former projection (C_3). In that comparison, the allocation disagreement is 0. However, the allocation probability disagreement is close to the value of C_R. The quantity disagreement of C_3 is not defined. At the same time, the quantity probability disagreement is clearly above the C_R value.

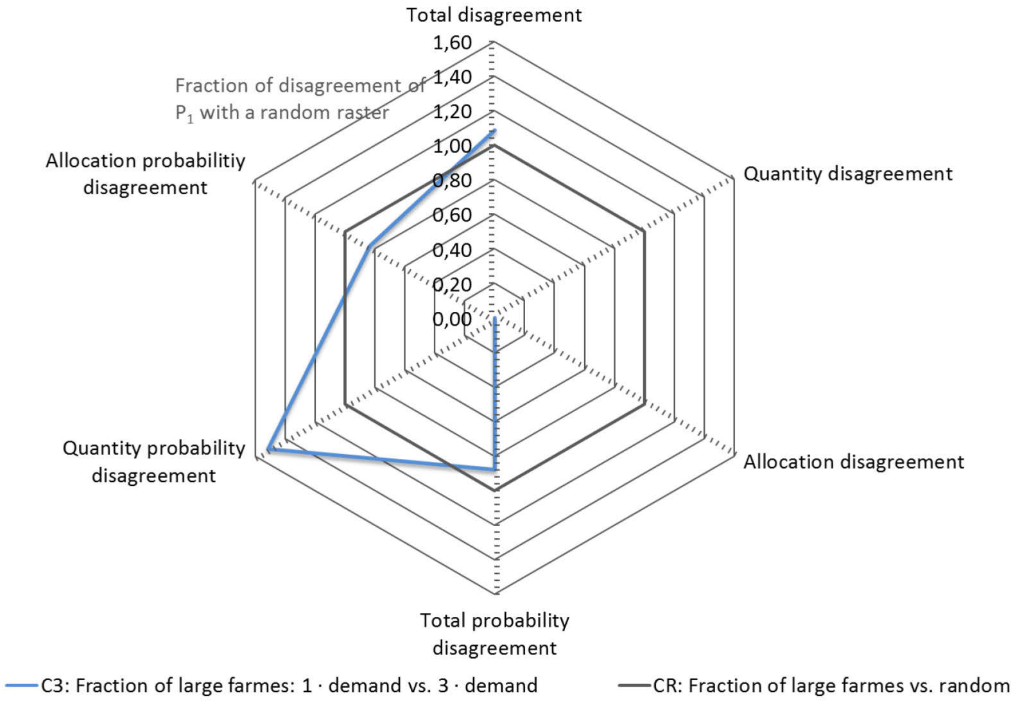


Figure IV-10: Comparison of land change projections with different demands of land change

4 Discussion

4.1 Binary or continuous?

In this article we propose the use of six different disagreement measures to compare two land change projections. We divided these measures into two different levels. Both of these levels include the two disagreement components of quantity and space. The third component is the sum of the first two disagreement components in both levels. Many studies have highlighted that the differentiation of the disagreement into quantity and spatial components is useful (e.g. Veldkamp and Lambin 2001; Pontius and Millones 2011) to identify the strengths and weaknesses of the land change model. A comparison of classified or binary future land change is the most common approach (Brown *et al.* 2013). In contrast, comparing land change probabilities is rarely applied (e.g. Foody 2006; Sangermano *et al.* 2012). However, is it beneficial to investigate the two mentioned levels of comparing classified and probability maps at the same time?

Our findings show differing results for the quantity and location disagreement depending on the disagreement level. The comparison “fraction of large farms” to “distance to populated places” has a higher disagreement than the comparison “fraction of large farms” to “elevation” when considering the binary disagreement level. The continuous disagreement level shows the opposite result. As mentioned in our introductory example, a large difference of the probabilities of 0.4 can have the same label in both binary land change maps and a small difference of the probabilities of 0.1 can lead to the same land change category in both projections. Both disagreement levels address different issues. Therefore, the correlation of these levels is below 1. Both levels give a certain degree of information.

On the one hand, classified maps are derived from the continuous model outputs. It is a process of aggregation and thus a loss of information. The results are especially dependent on the chosen threshold which is used to differentiate between change and persistence pixels (Lobo *et al.* 2008). For that reason, it makes sense to compare the original model outputs. On the other hand, we are interested if land change can be expected or not. If a certain level of suitability of a pixel is exceeded, this pixel will be converted into another class, no matter if the original continuous value was 0.5 or 0.9.

Comparing the classified land change maps is especially valuable when we are interested in real impacts. It is not crucial how suitable a pixel is, it is only important if the suitability value is below or above a defined threshold. In terms of probabilities, it is only important if the probability of land change is below, equal to or above 0.5. When comparing the classified maps, the predictive ability is evaluated.

Comparing the continuous probability maps is especially valuable when we are interested in comparing the tendencies to project land change. We know that every projection has a certain degree of uncertainty which is implied in the probabilities (Krüger and Lakes 2014). The probabilities of 0.1 and 0.4 lead to the same decision about projected land change; however, 0.1 has less uncertainty. The closer the value to 0.5, the more uncertainty is implied. We can interpret the comparison of continuous land change maps as the comparison of the binary maps weighted with uncertainty. More uncertain differences are weakened and less uncertain differences are highlighted in the calculation of the probability disagreement measures. Moreover, comparing the continuous maps gives evidence about the discriminative ability of a model, i.e. the distinction between change and persistence pixels.

4.2 What does similar mean?

One intention in the comparison of different projections is the investigation of the uncertainty of future projections. It is not possible to compare the model output with any true land change data until the projected future point in time has happened. Therefore, we compare two different projections, both of them based on realistic assumptions. No disagreement in the produced land change output maps would substantiate the expectation that there is little uncertainty in the modeled future land change. By comparison, considerable differences would be an indicator for high uncertainty in the modeled land change.

Given the presented disagreement dimensions in this article, we get a certain disagreement in any of these dimensions, e.g. 0.5. Is 0.5 a strong difference? To approach this question, we chose a random land change raster as a reference map. A random raster has no explanatory potential for a systematic land change process. At the same time, we assume a certain degree of explanatory potential for an applied land change model. The output of the land change model is a result of the given systematic input data. Thus, the agreement between the unsystematic random raster and the systematic modeled raster is only chance. The disagreement is high. Therefore, a disagreement of equal or more than the disagreement with the random raster is considerable. Some comparison results in this article have such a high disagreement value. In these cases, the uncertainty about future land change is high. In the case of a disagreement of 0 between the two projections, there is clearly no uncertainty about future land change given the two projections. However, how can we address the interval between 0 and 1 as a fraction of the disagreement with a random raster? Is it possible to set a threshold to differentiate minor differences from considerable differences?

There is no universally valid answer to this question. The meaning of similarity depends on the chosen research problem. A disagreement of 0.5 between two projections in the presented spider chart means that half of the disagreement which is termed considerable is substituted by agreement. Half of the uncertainty about future land change is still present. This statement helps to interpret different modeling results.

It is possible to use other reference comparisons such as the one used in this article. We proposed a comparison based on a random allocation of a given demand. This demand is additionally used for most of the presented land change projections. Therefore, the quantity disagreement is 0 in these cases. Alternatives would be to choose a purely random raster with randomly set proportions of the land change classes, a model which predicts "land persistence” at every pixel (Pontius *et al.* 2008), or a model which predicts deforestation near former deforested areas (Diniz *et al.* 2013).

4.3 Limitations and recommendations

One limitation of this study is the restriction to binary land change examples. However, these binary problems are widely applied in the land change community (e.g. Lakes *et al.* 2009). Moreover, a multi-class land change problem can be translated into several two-class land change problems. In this case, the uncertainty statement is valid for one land change transition instead of the whole case study problem.

Furthermore, the effect of the chosen storylines on the quantified uncertainty has to be considered. Two contrasting storylines will most likely lead to substantially different land change outputs. Following the logic of this study, the uncertainty about future land change would be very high. This conclusion is biased if at least one of the storylines is really unrealistic. Therefore, interpretations about the uncertainty depend on the reasonability of the underlying storylines.

In the land change modeling community there are two opposing research directions about how to use projections of future land change. Studies belonging to the first direction intend to reveal possible future developments without quantifying any uncertainty. The aim of these studies is to show what could happen if certain circumstances are given (e. g. a specific law of environmental protection would be enacted). Following Pontius and Neeti (2010) the results of these studies are frequently termed as scenarios which are mainly determined by the qualitative storylines. A quantification of uncertainty makes no sense in these cases. Our approach is useful for the second research direction. Therein, projections are based on a predictive model. Storylines are solely the frame of the projection and the shape of the projected land change is an output of the calibrated model.

5 Conclusions

Land change modelers would like to know if their models deliver useful results. Therefore, models are frequently validated with true reference data. If a land change model is used to project the quantity and location of future land change, this kind of reference does not exist.

In this article, we suggest the use of a second projection and compare these two projections. Given that both projections are based on reasonable assumptions, substantial differences imply a high uncertainty in the projection of future land change. In comparison, weak differences substantiate a lower degree of uncertainty. We propose the use of six different uncertainty measures to cover the disagreement of quantity versus the disagreement of space, and the disagreement of real land change implications versus the disagreement of the general land change tendency adjusted for uncertainty. Moreover, we present a useful reference comparison which helps to differentiate similar and different map pairs in the projection.

Acknowledgments and funding

We thank Dennis Funke for a lot of data processing work. We additionally thank Florian Gollnow for helpful discussions that helped to improve this manuscript.

The presented work is supported by Deutsche Forschungsgemeinschaft [GRK 1324/1, METRIK].

Chapter V:
Synthesis

1 Summary of the main contributions

Uncertainty is a crucial part of land change modeling. It emerges in every modeling step and propagates itself through the remaining framework of the land change model application: from the model calibration to the validation to the projection. This work addressed the distinct steps within this modeling process by means of two research objectives.

Research objective 1 was to develop an approach to systematically identify and analyze uncertainties in land change modeling. This thesis addressed this in the following way:

- This thesis identified Bayesian Belief Networks (BBNs) as a valuable modelling approach to involve an uncertainty investigation in land change modeling. Furthermore, uncertainty measures based on this modeling approach were created and adopted to address different uncertainty sources.

BBNs reflect uncertainties by means of probabilities. BBNs are a graphical representation of the dependencies between variables. Assume that a node “land change” is dependent on a node “population density” which was one result of an applied expert survey. Technically, a link between “population density” and “land change” is included. This link expresses the conditional dependency between the two variables. Every state of the “land change” node is conditionally dependent on the states of the “population density” node and therefore a specific probability is assigned. This probability reflects uncertainty and can be used to analyze the effect of uncertainty on the modeling outcome. A uniform probability distribution describes the highest possible uncertainty. In a two-classes-problem, the maximum uncertainty is a 0.5 probability for the first class and a 0.5 probability for the second class. In this thesis, the probability distribution was summarized using the Mutual information Criterion. This measure is further used to quantify the uncertainty of different modeling steps. Moreover, the measure of “uncertain node” was developed and derived from the model building process to address structural uncertainty. The Mutual information Criterion was used to assess the uncertainty in land change application before (e.g. Foody 2006); however, this study utilized the measure adjusted for different land change uncertainty sources within one application.

- The three eminent uncertainty sources connected to the modeling steps of model structure, variable selection and data preprocessing were comprehensively investigated.

Investigating different uncertainty sources in one case study is frequently performed by means of a sensitivity analysis which addresses the influence on the modeling outcome (Lilburne and Tarantola 2009; Alcamo *et al.* 2011). Most studies focus on exploring the output uncertainty when changing the input parameters under one specific model structure (Verburg *et al.* 2013). Therein lies one of the important added values of this dissertation's uncertainty methodology. The effects of different uncertainty sources can be analyzed in the light of different model structures. By means of the presented BBN approach, it is additionally possible to analyze the dependencies between different uncertainty sources within the model calibration process. In the Brazilian case study, only few variables were identified as uncertain nodes, i.e. the fundamental structure of the land change model is relatively certain. However, the individual influence varied, even in two learned models with the same structure. Therefore, model structure uncertainty is not negligible. This thesis further identified that most of the ability to reduce uncertainty was concentrated on only few variables. Moreover, a model with a subset of variables led to less remaining uncertainty about land change than a model with all available variables. This highlights the fact that a complex model is not the best choice in every case study. The same conclusion is valid with regard to data preprocessing uncertainty. The exemplarily used variant of disaggregated population density data led to more remaining uncertainty in the model.

- An approach was developed to analyze the effect of single uncertainty sources on the type of disagreement in the modeling outcome, i.e. a separation into an incorrect spatial allocation of land change pixels and an incorrect quantity of the land change class.

Bayesian Belief Networks were previously used to investigate the propagation of uncertainty with regard to a final land change outcome (Laskey *et al.* 2010). However, a consideration whether the uncertainty has predominantly spatial or quantitative effects on the land change projections is missing. The thesis closed this gap. First, it was identified that the amount of uncertainty dedicated to a specific source and its importance for the accuracy of the modeling results do not have to be connected in every case. Different uncertainty sources are interdependent and can mitigate or enhance each other through propagation from the source to the final land change map. Furthermore, for the case study this study identified that the ratio of the disagreement due to an incorrect quantity of land change and the disagreement due to an incorrect spatial localization was not dependent on the amount of uncertainty of the three analyzed sources. This is a case study specific result. It is possible that the ratio between spatial and quantity disagreement in the model output is sensitive to the amount of uncertainty of a certain input source. By means of the proposed uncertainty investigation, model result recipients get an impression about how model weaknesses can influence the reliability of the output.

The second research objective was to develop methods to quantify uncertainty in land change projections which differentiate between spatial and quantitative uncertainty. The following problems are solved in the context of this objective:

- Two new measures of uncertainty were created which separate the total uncertainty into quantity and spatial components.

These measures are based on probabilities and only need one land change map, given in its original continuous shape. When the original outputs of the model are propensities, which give no information about the quantity of change, it is possible to transfer this output into probabilities. The developed uncertainty measurement has similarities to some existing concepts. Van Vliet *et al.* (2013) used a fuzzy approach to include a degree of thematic and spatial uncertainty. Additionally, Pontius and Millones (2011) used separation of the disagreement into spatial and quantity disagreement as performed in this thesis. The benefit of the newly developed uncertainty measures is that they are calculated without reference data, and are therefore applicable in future time steps. The created measures help to identify if a future process is certain or uncertain. However, they cannot determine the goodness-of-fit of the land change projections. Even an absolutely certain land change model can be totally wrong.

- A framework on how to use the quantified uncertainty was developed to estimate the reliability of future land change projections.

The accuracy of future land change projections cannot be calculated. Some authors suggested estimating the reliability of land change predictions by extrapolating the accuracy decay in known time steps into the future (Pontius and Spencer 2005; Pontius and Neeti 2010). The problem with this approach is that unsystematic developments are not incorporated. Chapter (III) gives another possible solution to the dilemma described in the paragraph above. This dissertation proposes the quantification of the relationship between disagreement and uncertainty in known time steps when all measures are applicable. In cases when high uncertainty and high disagreement (and vice versa) are detected under the same circumstances (e.g. subregions, time intervals, external circumstances), it is reasonable to assume that this relationship is still valid in land change projections. Unsystematic changes are included in this approach when the calculated uncertainty reflects an unsystematic development over time. Nevertheless, projecting past relationships or trends into the future is not necessarily realistic in every case study (Müller *et al.* 2014). However, this is evidence for the reliability of the land change model for future projections.

- An additional approach was developed to quantify future uncertainty about land change based on a comparison of different land change projections.

Another possibility for addressing future uncertainty is given in chapter (IV) which is based on the comparison of two projections. Previous studies addressed uncertainty in future land change projections either without giving any quantitative information about their differences (e.g. Verburg *et al.* 2010), by calculating the different rates of change (Hoymann 2011; Rodríguez Eraso *et al.* 2013; Schmitz *et al.* 2014), or by measuring the differences of different classified projections (Sohl *et al.* 2012). However, Pontius and Neeti (2010) identified that future uncertainty can obscure a high fraction of differences. This means that the classified output maps can be substantially different due to different underlying storylines; however, classified pixels adjusted for uncertainty can be similar at the same time. Therefore, the methods in this thesis augment the previously applied approaches of investigating uncertainty of future land change. A six-dimensional comparison was developed which is summarized in a spider chart. This comparison was applied on different reasonable projections to outline an uncertainty range. Three known dimensions address the disagreement of the classified outputs and reflect the disagreement in the real implications of projected land change. The drawback of such disagreement measures is that they are dependent on a chosen threshold separating land use change from land use persistence (Lobo *et al.* 2008). Therefore, three dimensions newly developed in this thesis quantify the differences in the modeled probabilities. These dimensions reflect the differences in land change tendencies which are adjusted for the modeled uncertainty. This implies a comparison of the ability to discriminate between land change classes.

- The thesis gives guidance on how to interpret differences between two projections by means of useful reference comparisons.

Once the differences between two or more projections are detected, we have to decide if these differences are substantial and if the uncertainty about future land change processes is high. A detected disagreement should be related to the amount or variability of change in the study area (Brown *et al.* 2005; Diniz *et al.* 2013). For this purpose, a reference comparison was developed. This study proposes a random process without useful information about future land change. Differences between a projection and the random reference help to range differences between this particular projection and a second one. The spider chart with the six dimensions is adjusted to the reference comparison. A difference of one means that the difference is as high as with a random land change map. This is helpful when classifying similarity and difference. An alternative reference comparison is a projection which is based only on the distance to previously changed areas. There are other published studies which emphasize the importance of relating the measured uncertainty to the amount of uncertainty which can be expected (e.g. van Vliet *et al.* 2011). One added value of this dissertation is that the user is flexible in the choice of the reference comparison. Therefore, the proposed uncertainty measurement leaves room to include the requirements of the specific application.

2 Limitations and recommendations

The methods developed and applied in this thesis have limitations. No methodology is 100 % valid in every possible case study. However, the land change modeler must to be aware of the opportunities and challenges of the chosen approach. Therefore, potential difficulties are given along with suggestions about how to deal with them.

Concerning the first research objectives, some drawbacks of using BBNs have been discussed in previous studies. BBNs are graphical representations of the dependencies between input variables; the dependencies are quantified with conditional probability tables. These probability tables can become very large if one variable directly depends on a number of other variables. This challenge can be addressed by following the guidelines for constructing BBNs given in Marcot *et al.* (2006). Another concern regarding BBNs is the limited ability for representing complex processes. It is difficult to consider feedbacks and loops between variables or to incorporate a temporal or spatial dynamic (Uusitalo 2007). Nevertheless, BBNs are increasingly applied in environmental modeling, among others because of their explicit and intuitively accessible representation of uncertainty (Aguilera *et al.* 2011). Moreover, BBNs allow the integration of different knowledge sources and the straightforward assessment of what-if scenarios, e.g. what is the probability of land change if population density increases. The drawbacks and advantages of BBNs have to be weighed before applying this method. This thesis identified that BBNs are especially versatile when a comprehensive uncertainty investigation is needed. However, BBNs show their limitations in case studies where the entire complexity of land change processes should be represented in the model. One recommendation would therefore be to use BBNs in combination with other approaches. Complex presses can be represented in submodels which are integrated with a BBN. In this way, the strengths of different modeling approaches can be combined.

The measures which were developed in this thesis are related to a binary case study, i.e. the separation into land use change and land use persistence. However, several land change classes are possible within one application. For example, a change from forest to pasture or cropland is possible. This challenge can be addressed by breaking the multi-class problems down into several two-class problems. Every possible transition is transferred to the decision of either the change into one specific class or no change. The three-class example becomes a two-times-two-class application. Therefore, the developed uncertainty measures are still applicable. However, they have a modified meaning. The measures represent the spatial and quantitative uncertainty connected to one specific transition rather than the uncertainty of the whole application.

In this thesis, spatial and quantity uncertainties have been investigated. Moreover, other dimensions of uncertainty affect the reliability of land change modeling results. Temporal uncertainty, which is the uncertainty about the specific point in time when land change will occur, is one further dimension. It is possible that a model predicts land change in time step t_1 and no land change occurs in this time step, even when land changes shortly after t_1. This dimension could be a subject of future work. However, it can be reasoned that temporal uncertainty is already implicitly included in quantity uncertainty. Following the demand and spatial allocation modeling approach of this dissertation, a certain location can be estimated as land change when the probability of land change is higher as 1 minus demand, assuming that the demand is a function over time. For example, in most business-as-usual scenarios, the demand increases over time. The specification of the demand and the question when land will change are interrelated. Another uncertainty dimension is thematic uncertainty. It is certain that land change will occur; however, the specific transition in a multi-class application is uncertain. This type of uncertainty can be explored when transition specific uncertainties are calculated as recommended above. In cases where several transitions have a similar probability, thematic uncertainty is high.

This thesis investigated different uncertainty sources and applied a set of newly developed measures to quantify the effects of these sources. One essential part of the modeling strategy applied in this dissertation is the definition of the land change demand which itself is subject to uncertainty. One possibility is the estimation of the demand by means of global economic models, which are prone to the same uncertainty sources defined by Walker *et al.* (2003). An alternative approach is to define the demand based on storylines such as the IPCC storylines (Nakicenovic and Swart 2000). The quantification of land change based on qualitative storylines is however again a source of uncertainty. Different experts translate the storylines into a different amount of future land change (Verburg *et al.* 2013). Land change studies could address the issue by including different expert opinions or by requesting the level of confidence in the defined demand. Such a level of confidence could be included into the model, for example as an additional node in a BBN. Then different possible demands could be assigned a respective probability.

Major parts of this thesis dealt with the quantification of future uncertainty. In the land change modeling community there are two opposing views about how to address future uncertainty. Some researchers use scenarios to show plausible future developments given several underlying assumptions (Mancosu *et al.* 2015). This kind of analysis has the characteristic of an if-then consideration. Different scenarios should show the rage of future uncertainty without explicitly quantifying the uncertainty. This research stream considers future land change as strongly dependent on the underlying assumptions which are made for the modeling approach. Since most of the uncertainty is covered by these assumptions, it makes no sense to quantify the uncertainty of one scenario (Pontius and Neeti 2010). Another point of view is to categorize modeling results for future land change as a projection (Verstegen *et al.* 2015), as it has been done in this thesis. Given assumptions or storylines are only the foundation of one projection. The specific shape of the land change is the result of the predictive land change model. Such a model is calibrated with quantitative data and quantitatively applied in future time steps. Therefore, a deviation to the truth could be applied if the truth becomes reality. Until this will happen, the identification of the deviation to the unknown truth makes sense. That means, the quantification of uncertainty is useful in these cases.

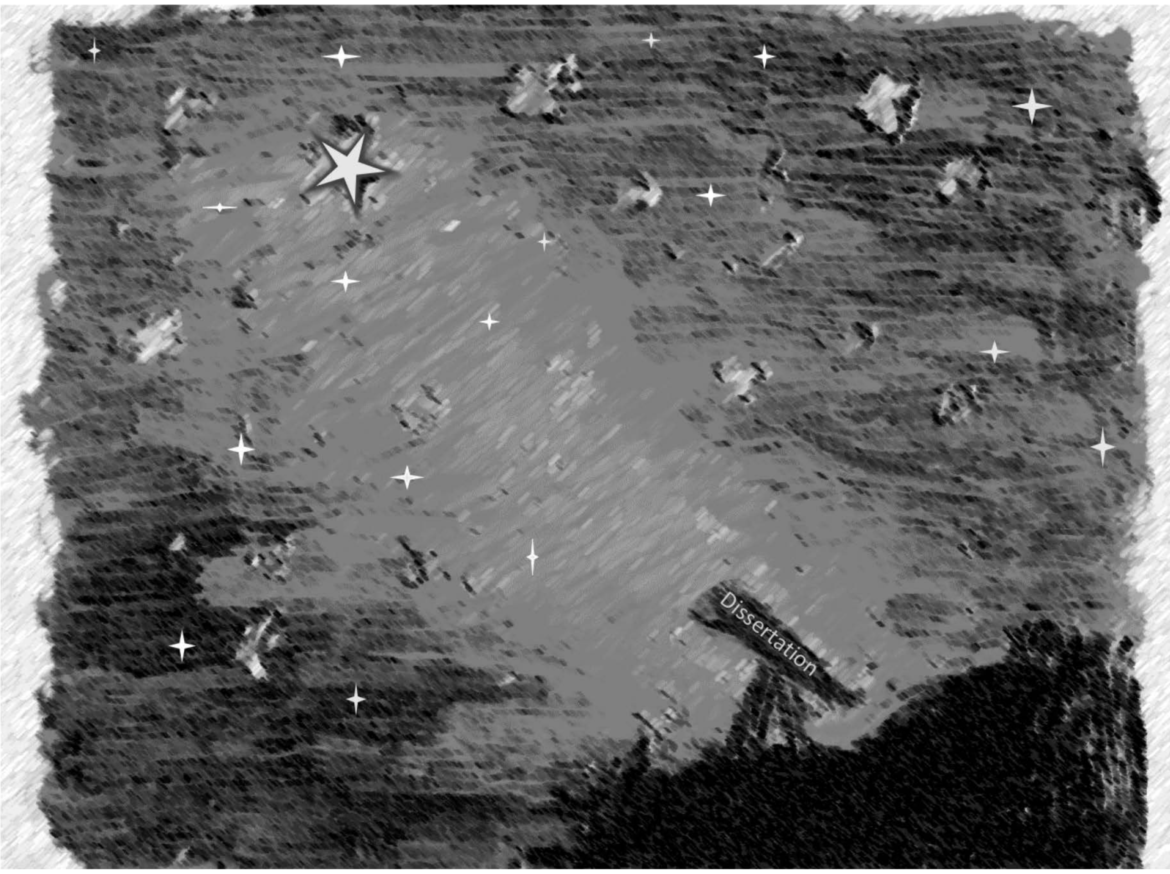
The methods of this thesis leave room for misinterpretations. When uncertainty is measured, nothing is said about the model's goodness-of-fit. A certain land change model can predict every pixel wrong and vice versa. Nevertheless, this thesis deduced that uncertainty can be an indicator for accuracy. It is comprehensible that highly uncertain environmental processes are difficult to predict. Therefore, the hypothesis at the beginning of every land change study could be that disagreement is dependent on uncertainty. This dependency can be analyzed empirically and subsequently used to reason about future disagreement. However, the additional assumption is implied that an identified relationship between uncertainty and disagreement is stationary in time. This assumption is not necessarily realistic. In their Amazon case study, Rosa et al. (2015) identified that the reliability of land change projections is strongly dependent on the chosen calibration time interval. The fact remains that future disagreement is not measurable. However, measured uncertainty can be a basis for estimating disagreement. This disagreement increases in time in most applications and modelers are aware that the reliability of land change predictions decreases in future time steps (Chaudhuri and Clarke 2014; Qiang and Lam 2015) though the decreasing rate is unknown. Depending on the application, there is a certain point in time when land change predictions no longer give useful information for decision-making. The approaches developed in this thesis could help to identify this unknown point in time.

Another misinterpretation can emerge when comparing the uncertainty of different applications. In situations with a lot of unsystematic variability, land change models are necessarily less certain in their prediction than in deterministic cases. Certainty is unequal to reliability in such investigations of different case studies. However, in an application of different model settings or modeling approaches within one case study, uncertainty gives evidence about reliability. Then, it is reasonable to conclude that a certain model is more reliable than an uncertain model. That is the procedure suggested in chapter four.

3 Conclusions

Land change modeling is crucial to understand the complex processes in the human-environment system. It gives important guidance for land use planning and decision-making. On the one hand, it is necessary that decision-makers have confidence in the utility of land change modeling outputs; on the other hand, they have to be aware of model weaknesses in order to avoid misinterpreting the results. Therefore, it is of utmost importance to give information about uncertainty within the modeling process and in the model results. This work presented a comprehensive approach about how to identify and evaluate the uncertainties in the model calibration. The approach is additionally a profound basis to assess the effects of different uncertainty sources on the reliability of the model's outcome. It turned out in this thesis that such effects are complex and hard to predict without a quantitative uncertainty assessment. Furthermore, two procedures to quantify uncertainty of land change projections in future time horizons were developed. Many land change applications express the expectation that uncertainty increases when going further into the future; however, this work fills the quantitative gap of these expectations. Moreover, the developed approaches allow the differentiation of uncertainty in spatial and quantitative components, which is an important asset in spatial applications such as land change modeling.

A dissertation, such as this one, is like a telescope focusing on a specific part of the dark sky. Within this dissertation, the focus was on uncertainties relying on probabilities and on a differentiation of these uncertainties into spatial and quantity uncertainty. That is only a tiny part of the widespread topic of uncertainty in spatial modeling, and it is even only a tiny part of the more specific topic of uncertainty in land change modeling. Nevertheless, I believe that I could find at least one bright star which brought some light into the dark sky. With a certain distance, one star looks quite similar to the uncountable number of neighboring stars…



© Jördis Krüger

„Das Interessante am Sternenhimmel sind nicht die Sterne, sondern die Zwischenräume.“

(Gerald Dunkl (*1959), österreichischer Psychologe und Aphoristiker)

References

Aalders, I., 2008. Modeling Land-Use Decision Behavior with Bayesian Belief Networks. Ecology and society a journal of integrative science for resilience and sustainability, 13 (1), 16–37. Available from: http://www.ecologyandsociety.org/vol13/iss1/art16/ [Accessed 5 September 2013].

Aerts, J.C.J.H., Goodchild, M.F., and Heuvelink, G.B.M., 2003. Accounting for Spatial Uncertainty in Optimization with Spatial Decision Support Systems. Transactions in GIS, 7 (2), 211–230. DOI: http://dx.doi.org/10.1111/1467-9671.00141.

Aguiar, A.P.D., Câmara, G., and Escada, M.I.S., 2007. Spatial statistical analysis of land-use determinants in the Brazilian Amazonia: Exploring intra-regional heterogeneity. Ecological Modelling, 209 (2–4), 169–188. DOI: http://dx.doi.org/10.1016/j.ecolmodel.2007.06.019.

Aguiar, A.P.D., et al., 2012. Modeling the spatial and temporal heterogeneity of deforestation - driven carbon emissions: the INPE - EM framework applied to the Brazilian Amazon. Global Change Biology, 18 (11), 3346 - 3366. DOI: http://dx.doi.org/10.1111/j.1365-2486.2012.02782.x.

Aguilera, P.A., et al., 2011. Bayesian networks in environmental modelling. Environmental Modelling & Software, 26 (12), 1376–1388. DOI: http://dx.doi.org/10.1016/j.envsoft.2011.06.004.

Alcamo, J., et al., 2011. Evaluation of an integrated land use change model including a scenario analysis of land use change for continental Africa. Environmental Modelling & Software, 26 (8), 1017–1027. DOI: http://dx.doi.org/10.1016/j.envsoft.2011.03.002.

Aldwaik, S.Z., and Pontius, R.G., 2013. Map errors that could account for deviations from a uniform intensity of land change. International Journal of Geographical Information Science, 1–23. DOI: http://dx.doi.org/10.1080/13658816.2013.787618.

Araujo, C., et al., 2009. Property rights and deforestation in the Brazilian Amazon. Ecological Economics, 68 (8–9), 2461–2468. DOI: http://dx.doi.org/10.1016/j.ecolecon.2008.12.015.

Arima, E.Y., et al., 2011. Statistical confirmation of indirect land use change in the Brazilian Amazon. Environmental Research Letters, 6 (2), 24010. Available from: http://stacks.iop.org/1748-9326/6/i=2/a=024010 [Accessed 5 September 2013].

Arsanjani, J.J., et al., 2013. Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion. International Journal of Applied Earth Observation and Geoinformation, 21, 265–275. DOI: http://dx.doi.org/10.1016/j.jag.2011.12.014.

Ascough II, J.C., et al., 2008. Future research challenges for incorporation of uncertainty in environmental and ecological decision-making. The Importance of Uncertainty and Sensitivity Analysis in Process-based Models of Carbon and Nitrogen Cycling in Terrestrial Ecosystems with Particular Emphasis on Forest Ecosystems Selected Papers from a Workshop Organized by the International Society for Ecological Modelling (ISEM) at the Third Biennal Meeting of the International Environmental Modelling and Software Society (IEMSS) in Burlington, Vermont, USA, August 9-13, 2006. Ecological Modelling, 219 (3–4), 383–399. DOI: http://dx.doi.org/10.1016/j.ecolmodel.2008.07.015.

Asselen, S., and Verburg, P.H., 2013. Land cover change or land - use intensification: simulating land system change with a global - scale land change model. Global Change Biology, 19 (12), 3648 - 3667. DOI: http://dx.doi.org/10.1111/gcb.12331.

Barona, E., et al., 2010. The role of pasture and soybean in deforestation of the Brazilian Amazon. Environmental Research Letters, 5 (2), 24002.

Bashari, H., Smith, C., and Bosch, O.J.H., 2008. Developing decision support tools for rangeland management by combining state and transition models and Bayesian belief networks. Agricultural Systems, 99 (1), 23–34. DOI: http://dx.doi.org/10.1016/j.agsy.2008.09.003.

Bastin, L., et al., 2012 Managing uncertainty in integrated environmental modelling: The UncertWeb framework. Environmental Modelling & Software, 2012 (0). DOI: http://dx.doi.org/10.1016/j.envsoft.2012.02.008.

Batisani, N., and Yarnal, B., 2009. Uncertainty awareness in urban sprawl simulations: Lessons from a small US metropolitan region. Land Use Policy, 26 (2), 178–185. DOI: http://dx.doi.org/10.1016/j.landusepol.2008.01.013.

Börner, J., et al., 2015. Post-Crackdown Effectiveness of Field-Based Forest Law Enforcement in the Brazilian Amazon. DOI: http://dx.doi.org/10.1371/journal.pone.0121544.

Boucher, D., Roquemore, S., and Fitzhugh, E., 2013. Brazil's success in reducing deforestation. Tropical Conservation Science, 6 (3), 426–445. Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-84882731426&partnerID=40&md5=06c17c8f1d73618671a523c406c3b9f4 [Accessed 15 July 2015].

Boutilier, C., Friedman, N., and Goldszmidt, M., eds., 1996. Context-specific independence in Bayesian networks: Morgan Kaufmann Publishers Inc.

Bromley, J., 2005. Guidelines for the use of Bayesian networks as a participatory tool for Water Resource Management.

Brown, D.G., et al., 2005. Path dependence and the validation of agent - based spatial models of land use. International Journal of Geographical Information Science. International Journal of Geographical Information Science, 19 (2), 153 - 174. DOI: http://dx.doi.org/10.1080/13658810410001713399.

Brown, D.G., et al., 2013. Opportunities to improve impact, integration, and evaluation of land change models. Current Opinion in Environmental Sustainability, 5 (5), 452–457. DOI: http://dx.doi.org/10.1016/j.cosust.2013.07.012.

Brown, J.D., and Heuvelink, G.B.M., 2007. The Data Uncertainty Engine (DUE): A software tool for assessing and simulating uncertain environmental variables. Computers & Geosciences, 33 (2), 172–190. DOI: http://dx.doi.org/10.1016/j.cageo.2006.06.015.

Celentano, D., et al., 2012. Welfare Outcomes and the Advance of the Deforestation Frontier in the Brazilian Amazon. World Development, 40 (4), 850–864. DOI: http://dx.doi.org/10.1016/j.worlddev.2011.09.002.

Chaudhuri, G., and Clarke, K.C., 2014. Temporal accuracy in urban growth forecasting: A study using the SLEUTH model. Transactions in GIS, 18 (2), 302–320. DOI: http://dx.doi.org/10.1111/tgis.12047.

Chen, H., and Pontius, R.G., 2010. Diagnostic tools to evaluate a spatial land change projection along a gradient of an explanatory variable. Landscape Ecology, 25 (9), 1319-1331. DOI: http://dx.doi.org/10.1007/s10980-010-9519-5.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and psychological measurement, 20 (1), 37–46.

Comber, A., et al., 2012. Spatial analysis of remote sensing image classification accuracy. Remote Sensing of Environment, 127, 237–246. DOI: http://dx.doi.org/10.1016/j.rse.2012.09.005.

Comber, A.J., 2013. Geographically weighted methods for estimating local surfaces of overall, user and producer accuracies. Remote Sensing Letters, 4 (4), 373–380. DOI: http://dx.doi.org/10.1080/2150704X.2012.736694.

Coy, M., and Klingler, M., 2008. Pionierfronten im brasilianischen Amazonien zwischen alten Problemen und neuen Dynamiken. Das Beispiel des „Entwicklungskorridors“ Cuiabá (Mato Grosso) – Santarém (Pará). Innsbrucker Geographische Gesellschaft: Innsbrucker Jahresbericht 2008-2010 (10), 109–129.

Crosetto, M., and Tarantola, S., 2001. Uncertainty and sensitivity analysis: tools for GIS-based model implementation. International Journal of Geographical Information Science. International Journal of Geographical Information Science, 15 (5), 415–437. DOI: http://dx.doi.org/10.1080/13658810110053125..

Crosetto, M., Moreno Ruiz, J.A., and Crippa, B., 2001. Uncertainty propagation in models driven by remotely sensed data. Remote Sensing of Environment, 76 (3), 373–385. DOI: http://dx.doi.org/10.1016/S0034-4257(01)00184-5.

de Espindola, G.M., et al., 2012. Agricultural land use dynamics in the Brazilian Amazon based on remote sensing and census data. Applied Geography (32), 240–252. DOI: http://dx.doi.org/10.1016/j.apgeog.2011.04.003.

de Fries, R., et al., 2013. Export-oriented deforestation in Mato Grosso: harbinger or exception for other tropical forests? Philosophical transactions of the royal society B: biological sciences, 368 (1619), 20120173. DOI: http://dx.doi.org/10.1098/rstb.2012.0173.

Diniz, L.G., et al., eds., 2013. Measuring Allocation Errors in Land Change Models in Amazonia. Proceedings of XIV GEOINFO, November 24-27, 2013, Campos do Jordao, Brazil.

Duespohl, M., Frank, S., and Doell, P., 2012. A review of Bayesian networks as a participatory modeling approach in support of sustainable environmental management. Journal of Sustainable Development, 5 (12), p1. DOI: http://dx.doi.org/10.5539/jsd.v5n12p1.

Eva, H.D., et al., 2002. A vegetation map of South America.

Fearnside, P.M., 2008. Amazonia and deforestation. In: A.S. Goudie and D.J. Cuff, eds. The Oxford Companion to Global Change. New York, USA: Oxford University Press, 21–27.

Fenton, N.E., and Neil, M.D., 2012. Risk Assessment and Decision Analysis with Bayesian Networks: CRC Press.

Foley, J.A., et al., 2005. Global Consequences of Land Use. Science, 309 (5734), 570–574. DOI: http://dx.doi.org/10.1126/science.1111772.

Food and Agriculture Organization of the United Nations (FAO), 2015. AO/UNESCO Soil Map of the World. Available from: http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/faounesco-soil-map-of-the-world/en/ [Accessed 20 September 2015].

Foody, G.M., 2005. Local characterization of thematic classification accuracy through spatially constrained confusion matrices. International Journal of Remote Sensing, 26 (6), 1217–1228. DOI: http://dx.doi.org/10.1080/01431160512331326521.

Foody, G.M., 2006. What is the difference between two maps? A remote senser's view. Journal of Geographical Systems, 8 (2), 119–130. DOI: http://dx.doi.org/10.1007/s10109-006-0023-z.

Fotheringham, A.S., Brunsdon, C., and Charlton, M., 2002. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. 1st ed. New York, USA: Wiley.

Gallego, F.J., 2010. A population density grid of the European Union. Population and Environment, 31 (6), 460–473. DOI: http://dx.doi.org/10.1007/s11111-010-0108-y.

Geist, H.J., and Lambin, E.F., 2002. Proximate Causes and Underlying Driving Forces of Tropical Deforestation. BioScience. BioScience, 52 (2), 143–150. DOI: http://dx.doi.org/10.1641/0006-3568(2002)052[0143:PCAUDF]2.0.CO;2.

Glasziou, P., and Hilden, J., 1989. Test selection measures. Medical Decision Making, 9 (2), 133–141. DOI: http://dx.doi.org/10.1177/0272989X8900900208.

Goerlich, F.J., and Cantarino, I., 2013. A population density grid for Spain. International Journal of Geographical Information Science, 1–17. DOI: http://dx.doi.org/10.1080/13658816.2013.799283.

Goldewijk, K.K., and Verburg, P.H., 2013. Uncertainties in global-scale reconstructions of historical land use: an illustration using the HYDE data set. Landscape Ecology, 28 (5), 861–877. DOI: http://dx.doi.org/10.1007/s10980-013-9877-x.

Gollnow, F., and Lakes, T., 2014. Policy change, land use, and agriculture: The case of soy production and cattle ranching in Brazil, 2001–2012. Applied Geography, 55, 203–211. Doi: http://dx.doi.org/10.1016/j.apgeog.2014.09.003.

Gopal, S. 2009. Error in GIS (propagation and modeling). In: Kitchen, R. and Thrift, N. (Eds) International Encyclopedia of Human Geography, Elsevier, 586-594.

Guitierrez-Velez, V.H., and Pontius, R.G., 2012. Influence of carbon mapping and land change modelling on the prediction of carbon emissions from deforestation. Environmental Conservation, 39 (04), 325–336. DOI: http://dx.doi.org/10.1017/S0376892912000173.

Hagen, A., 2003. Fuzzy set approach to assessing similarity of categorical maps. International Journal of Geographical Information Science, 17 (3), 235–249. DOI: http://dx.doi.org/10.1080/13658810210157822.

Hagen-Zanker, A., 2009. An improved Fuzzy Kappa statistic that accounts for spatial autocorrelation. International Journal of Geographical Information Science, 23 (1), 61–73. DOI: http://dx.doi.org/10.1080/13658810802570317.

Hagen-Zanker, A., and Martens, P., 2008. Map comparison methods for comprehensive assessment of geosimulation models. In: . Computational Science and Its Applications–ICCSA 2008: Springer, 194–209. DOI: http://dx.doi.org/10.1007/978-3-540-69839-5_15.

Hagos, S., et al., 2014. Assessment of uncertainties in the response of the African monsoon precipitation to land use change simulated by a regional model. Climate Dynamics, 43 (9-10), 2765–2775. DOI: http://dx.doi.org/10.1007/s00382-014-2092-x.

Hausman, C., 2012. Biofuels and Land Use Change: Sugarcane and Soybean Acreage Response in Brazil. Environmental and Resource Economics, 51 (2), 163–187. DOI: http://dx.doi.org/10.1007/s10640-011-9493-7.

Heuvelink, G.B.M., Burrough, P.A., and Stein, A., 1989. Propagation of errors in spatial modelling with GIS. International Journal of Geographical Information Systems. International Journal of Geographical Information Systems, 3 (4), 303–322. DOI: http://dx.doi.org/10.1080/02693798908941518.

Hilferink, M., and Rietveld, P., 1999. Land Use Scanner: An integrated GIS based model for long term projections of land use in urban and rural areas. Journal of Geographical Systems, 1 (2), 155–177. DOI: http://dx.doi.org/10.1007/s101090050010.

Hoymann, J., 2011. Accelerating urban sprawl in depopulating regions: a scenario analysis for the Elbe River Basin. Regional Environmental Change, 11 (1), 73–86. DOI: http://dx.doi.org/10.1007/s10113-010-0120-x.

Instituto Brasileiro de Geografia e Estatística (IBGE), 2013. Estimativas populacionais para os municípios brasileiros em 01.07.2013. Available from: http://www.ibge.gov.br/home/estatistica/populacao/estimativa2013/estimativa_dou.shtm [Accessed 20 September 2015].

Insituto Nacional de Pesquisas Espaciais (INPE), 2015. Taxas anuais do desmatamento - 1988 até 2014 [online]. Available from: http://www.obt.inpe.br/prodes/prodes_1988_2014.htm [Accessed 20 September 2015].

Instituto Nacional de Pesquisas Espaciais, 2013. PRODES Digital [online]. Available from: http://www.dpi.inpe.br/prodesdigital/prodes.php [Accessed 10 July 2014].

Jantz, C., Drzyzga, S., and Maret, M., 2014. Calibrating and Validating a Simulation Model to Identify Drivers of Urban Land Cover Change in the Baltimore, MD Metropolitan Region. Land, 3 (3), 1158–1179. DOI: http://dx.doi.org/10.3390/land3031158.

Kim, O.S., 2010. An assessment of deforestation models for reducing emissions from deforestation and forest degradation (REDD). Transactions in GIS, 14 (5), 631–654. DOI: http://dx.doi.org/10.1111/j.1467-9671.2010.01227.x.

Kocabas, V., and Dragicevic, S., 2013. Bayesian networks and agent-based modeling approach for urban land-use and population density change: a BNAS model. Journal of Geographical Systems, 15 (4), 403–426. DOI: http://dx.doi.org/10.1007/s10109-012-0171-2.

Krayer von Krauss, M.P., et al., 2006. Response to “To what extent, and how, might uncertainty be defined” by Norton, Brown, and Mysiak. The Integrated Assessment Journal-Bridging Sciences & Policy, 6 (1), 89–94.

Krüger, C., and Lakes, T., 2014. Bayesian belief networks as a versatile method for assessing uncertainty in land-change modeling. International Journal of Geographical Information Science, 29 (1), 111–131. DOI: http://dx.doi.org/10.1080/13658816.2014.949265.

Krüger, C., and Lakes, T, forthcoming. Revealing Uncertainties in Land Change Modeling Using Probabilities. Transactions in GIS.

Lakes, T., Müller, D., and Krüger, C., 2009. Cropland change in southern Romania: a comparison of logistic regressions and artificial neural networks. Landscape Ecology, 24 (9), 1195–1206. DOI: http://dx.doi.org/10.1007/s10980-009-9404-2.

Lambin, E.F., and Geist, H.J., eds., 2006. Land-Use and Land-Cover Change: Local Processes and Global Impacts. Berlin Heidelberg: Springer.

Laskey, K.B., Wright, E.J., and da Costa, P.C.G., 2010. Envisioning uncertainty in geospatial information. Bayesian Model Views. International Journal of Approximate Reasoning, 51 (2), 209–223. DOI: http://dx.doi.org/10.1016/j.ijar.2009.05.011.

Laurance, W.F., et al., 2001. The future of the Brazilian Amazon. Science, 291 (5506), 988. DOI: http://dx.doi.org/10.1126/science.291.5503.438.

Laurance, W.F., et al., 2002. Predictors of deforestation in the Brazilian Amazon. Journal of Biogeography, 29 (5-6), 737–748. DOI: http://dx.doi.org/10.1046/j.1365-2699.2002.00721.x.

Lauritzen, S.L., and Jensen, F., 2001. Stable local computation with conditional Gaussian distributions. Statistics and Computing, 11 (2), 191–203. DOI: http://dx.doi.org/10.1023/A:1008935617754.

Leyk, S., Boesch, R., and Weibel, R., 2005. A Conceptual Framework for Uncertainty Investigation in Map-based Land Cover Change Modelling. Transactions in GIS, 9 (3), 291–322. DOI: http://dx.doi.org/10.1111/j.1467-9671.2005.00220.x.

Lilburne, L., and Tarantola, S., 2009. Sensitivity analysis of spatial models. International Journal of Geographical Information Science. International Journal of Geographical Information Science, 23 (2), 151–168. DOI: http://dx.doi.org/10.1080/13658810802094995.

Lobo, J.M., Jiménez-Valverde, A., and Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. Global ecology and Biogeography, 17 (2), 145–151. DOI: http://dx.doi.org/10.1111/j.1466-8238.2007.00358.x.

Magliocca, N.R., and Ellis, E.C., 2013. Using Pattern‐oriented Modeling (POM) to Cope with Uncertainty in Multi‐scale Agent‐based Models of Land Change. Transactions in GIS, 17 (6), 883–900. DOI: http://dx.doi.org/10.1111/tgis.12012.

Malizia, N., 2013. The Effect of Data Inaccuracy on Tests of Space‐Time Interaction. Transactions in GIS, 17 (3), 426–451. DOI: http://dx.doi.org/10.1111/j.1467-9671.2012.01350.x.

Mancosu, E., et al., 2015. Future land-use change scenarios for the Black Sea catchment. Environmental Science & Policy, 46, 26–36. DOI: http://dx.doi.org/10.1016/j.envsci.2014.02.008.

Marcot, B.G., et al., 2006. Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. Canadian Journal of Forest Research, 36 (12), 3063–3074. DOI: http://dx.doi.org/10.1139/x06-135.

Margaritis, D., 2003. Learning Bayesian network model structure from data: University of Pittsburgh.

Mas, J.-F., et al., 2013. A suite of tools for ROC analysis of spatial models. ISPRS International Journal of Geo-Information, 2 (3), 869–887. DOI: http://dx.doi.org/10.3390/ijgi2030869.

Mas, J.-F., et al., 2014. Inductive pattern-based land use/cover change models: A comparison of four software packages. Environmental Modelling & Software, 51, 94–111. DOI: http://dx.doi.org/10.1016/j.envsoft.2013.09.010.

Mc Closkey, J.T., Lilieholm, R.J., and Cronan, C., 2011. Using Bayesian belief networks to identify potential compatibilities and conflicts between development and landscape conservation. Landscape and Urban Planning, 101 (2), 190–203. DOI: http://dx.doi.org/10.1016/j.landurbplan.2011.02.011.

Messina, J.P., et al., 2008. Complex systems models and the management of error and uncertainty. Journal of Land Use Science, 3 (1), 11–25. DOI: http://dx.doi.org/10.1080/17474230802047989.

Meyfroidt, P., et al., 2013. Globalization of land use: distant drivers of land change and geographic displacement of land use. Current Opinion in Environmental Sustainability, 5 (5), 438–444. DOI: http://dx.doi.org/10.1016/j.cosust.2013.04.003.

Müller, D., et al., 2014. Regime shifts limit the predictability of land-system change. Global Environmental Change, 28, 75–83. DOI: http://dx.doi.org/10.1016/j.gloenvcha.2014.06.003.

Müller, R., et al., 2012. Proximate causes of deforestation in the Bolivian lowlands: an analysis of spatial dynamics. Regional Environmental Change, 12 (3), 445–459. DOI: http://dx.doi.org/10.1007/s10113-011-0259-0.

Nakicenovic, N., and Swart, R., 2000. Special report on emissions scenarios. Special Report on Emissions Scenarios, Edited by Nebojsa Nakicenovic and Robert Swart, pp. 612. ISBN 0521804930. Cambridge, UK: Cambridge University Press, July 2000., 1.

National Geophysical Data Center, 2013. Nighttime Lights Time Series [online]. Available from: http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html [Accessed 9 September 2013].

Neapolitan, R.E., 2004. Learning Bayesian Networks. Chicago, Illinois: Prentice Hall.

Olofsson, P., et al., 2013. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. Remote Sensing of Environment, 129, 122–131. DOI: http://dx.doi.org/10.1016/j.rse.2012.10.031.

Olofsson, P., et al., 2014. Good practices for estimating area and assessing accuracy of land change. Remote Sensing of Environment, 148, 42–57. DOI: http://dx.doi.org/10.1016/j.rse.2014.02.015.

Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems. Networks of Plausible Inference. San Francisco: Morgan Kaufman Publishers.

Pebesma, E.J., Jong, K. de, and Briggs, D., 2007. Interactive visualization of uncertain spatial and spatio - temporal data under different scenarios: an air quality example. International Journal of Geographical Information Science, 21 (5), 515 - 527. DOI: http://dx.doi.org/10.1080/13658810601064009.

Peel, M.C., Finlayson, B.L., and McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. Hydrology and Earth System Sciences Discussions Discussions, 4 (2), 439–473.

Peter C, et al., 2009. Applying Bayesian modelling to assess climate change effects on biofuel production. Clim Res, 40 (2-3), 249–260.

Pijanowski, B.C., et al., 2002. Using neural networks and GIS to forecast land use changes: a Land Transformation Model. Computers, Environment and Urban Systems, 26 (6), 553–575. DOI: http://dx.doi.org/10.1016/S0198-9715(01)00015-1.

Pollino, C.A., et al., 2007. Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment. Bayesian networks in water resource modelling and management. Environmental Modelling & Software, 22 (8), 1140–1152. DOI: http://dx.doi.org/10.1016/j.envsoft.2006.03.006.

Pontius, R.G., 2000. Quantification error versus location error in comparison of categorical maps. Photogrammetric Engineering and Remote Sensing, 66 (8), 1011–1016.

Pontius, R.G., 2002. Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. Photogrammetric Engineering and Remote Sensing, 68 (10), 1041–1050.

Pontius, R.G., and Batchu, K., 2003. Using the relative operating characteristic to quantify certainty in prediction of location of land cover change in India. Transactions in GIS, 7 (4), 467–484.

Pontius, R.G., and Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. International Journal of Remote Sensing. International Journal of Remote Sensing, 32 (15), 4407–4429. DOI: http://dx.doi.org/10.1080/01431161.2011.552923.

Pontius, R.G., and Neeti, N., 2010. Uncertainty in the difference between maps of future land change scenarios. Sustainability Science, 5 (1), 39–50. DOI: http://dx.doi.org/10.1007/s11625-009-0095-z.

Pontius, R.G., and Parmentier, B., 2014. Recommendations for using the relative operating characteristic (ROC). Landscape Ecology, 29 (3), 367–382. DOI: http://dx.doi.org/10.1007/s10980-013-9984-8.

Pontius, R.G., and Spencer, J., 2005. Uncertainty in extrapolations of predictive land-change models. Environment and Planning B: Planning and Design (32), 211–230. DOI: http://dx.doi.org/10.1068/b31152

Pontius, R.G., Cornell, J.D., and Hall, C.A.S., 2001. Modeling the spatial pattern of land-use change with GEOMOD2: application and validation for Costa Rica. Agriculture, Ecosystems & Environment, 85 (1), 191–203. DOI: http://dx.doi.org/10.1016/S0167-8809(01)00183-9.

Pontius, R.G., et al., 2007. Accuracy assessment for a simulation model of Amazonian deforestation. Annals of the Association of American Geographers, 97 (4), 677–695. DOI: http://dx.doi.org/10.1111/j.1467-8306.2007.00577.x.

Pontius, R.G., et al., 2008. Comparing the input, output, and validation maps for several models of land change. The Annals of Regional Science, 42 (1), 11–37. DOI: http://dx.doi.org/10.1007/s00168-007-0138-2.

Pontius, R.G., et al., 2013. Design and interpretation of intensity analysis illustrated by land change in Central Kalimantan, Indonesia. Land, 2 (3), 351–369. DOI: http://dx.doi.org/10.3390/land2030351.

Qiang, Y., and Lam, N.S.N., 2015. Modeling land use and land cover changes in a vulnerable coastal region using artificial neural networks and cellular automata. Environmental monitoring and assessment, 187 (3), 1–16. DOI: http://dx.doi.org/10.1007/s10661-015-4298-8.

Refsgaard, J.C., et al., 2006. A framework for dealing with uncertainty due to model structure error. Advances in Water Resources, 29 (11), 1586–1597. DOI: http://dx.doi.org/10.1016/j.advwatres.2005.11.013.

Refsgaard, J.C., et al., 2007. Uncertainty in the environmental modelling process – A framework and guidance. Environmental Modelling & Software, 22 (11), 1543–1556. DOI: http://dx.doi.org/10.1016/j.envsoft.2007.02.004.

Richards, P.D., Walker, R.T., and Arima, E.Y., 2014. Spatially complex land change: The Indirect effect of Brazil's agricultural sector on land use in Amazonia. Global Environmental Change, 29, 1–9. DOI: http://dx.doi.org/10.1016/j.gloenvcha.2014.06.011.

Rodríguez Eraso, N., Armenteras-Pascual, D., and Alumbreros, J.R., 2013. Land use and land cover change in the Colombian Andes: dynamics and future scenarios. Journal of Land Use Science, 8 (2), 154–174. DOI: http://dx.doi.org/10.1080/1747423X.2011.650228.

Rosa, I.M.D., et al., 2015. Modelling land cover change in the Brazilian Amazon: temporal changes in drivers and calibration issues. Regional Environmental Change, 15 (1), 123–137. DOI: http://dx.doi.org/10.1007/s10113-014-0614-z.

Sangermano, F., Eastman, J.R., and Zhu, H., 2010. Similarity Weighted Instance - based Learning for the Generation of Transition Potentials in Land Use Change Modeling. Transactions in GIS, 14 (5), 569 - 580. DOI: http://dx.doi.org/10.1111/j.1467-9671.2010.01226.x.

Sangermano, F., Toledano, J., and Eastman, J.R., 2012. Land cover change in the Bolivian Amazon and its implications for REDD+ and endemic biodiversity. Landscape Ecology, 27 (4), 571–584. DOI: http://dx.doi.org/10.1007/s10980-012-9710-y.

Schaldach, R., et al., 2011. An integrated approach to modelling land-use change on continental and global scales. Environmental Modelling & Software, 26 (8), 1041–1051. DOI: http://dx.doi.org/10.1016/j.envsoft.2011.02.013.

Schaldach, R., et al., 2013. Model-based analysis of the environmental impacts of grazing management on Eastern Mediterranean ecosystems in Jordan. Journal of environmental management, 127, S84. DOI: http://dx.doi.org/10.1016/j.jenvman.2012.11.024.

Schmitz, C., et al., 2014. Land - use change trajectories up to 2050: insights from a global agro - economic model comparison. Agricultural Economics, 45 (1), 69 - 84. DOI: http://dx.doi.org/10.1111/agec.12090.

Scutari, M., 2010. Learning Bayesian Networks with the bnlearn R Package. Journal of Statistical Software, 35 (3), 1–22. Available from: http://www.jstatsoft.org/v35/i03 [Accessed 5 September 2013].

Shannon, C.E., and Weaver, W., 1949. The mathematical theory of communication: University of Illinois Press.

Sigel, K., Klauer, B., and Pahl-Wostl, C., 2010. Conceptualising uncertainty in environmental decision-making: The example of the EU water framework directive. Ecological Economics, 69 (3), 502–510. DOI: http://dx.doi.org/10.1016/j.ecolecon.2009.11.012.

Sloan, S., and Pelletier, J., 2012. How accurately may we project tropical forest-cover change? A validation of a forward-looking baseline for REDD. Global Environmental Change, 22 (2), 440–453. DOI: http://dx.doi.org/10.1016/j.gloenvcha.2012.02.001.

Soares-Filho, B.S., et al., 2006. Modelling conservation in the Amazon basin. Nature, 440 (7083), 520–523. DOI: http://dx.doi.org/10.1038/nature04389.

Sohl, T.L., et al., 2012. Spatially explicit land-use and land-cover scenarios for the Great Plains of the United States. Agriculture, Ecosystems & Environment, 153, 1–15. DOI: http://dx.doi.org/10.1016/j.agee.2012.02.019.

Stassopoulou, A., Petrou, M., and Kittler, J., 1998. Application of a Bayesian network in a GIS based decision making system. International Journal of Geographical Information Science, 12 (1), 23–46. DOI: http://dx.doi.org/10.1080/136588198241996.

Sun, Z., and Müller, D., 2013. A framework for modeling payments for ecosystem services with agent-based models, Bayesian belief networks and opinion dynamics models. Environmental Modelling & Software, 45, 15–28. DOI: http://dx.doi.org/10.1016/j.envsoft.2012.06.007.

Tayyebi, A., and Pijanowski, B.C., 2014. Modeling multiple land use changes using ANN, CART and MARS: Comparing tradeoffs in goodness of fit and explanatory power of data mining tools. International Journal of Applied Earth Observation and Geoinformation, 28, 102–116. DOI: http://dx.doi.org/10.1016/j.jag.2013.11.008.

Tayyebi, A., et al., 2014. Comparing three global parametric and local non-parametric models to simulate land use change in diverse areas of the world. Environmental Modelling & Software, 59, 202–221. DOI: http://dx.doi.org/10.1016/j.envsoft.2014.05.022.

Tayyebi, A.H., Tayyebi, A., and Khanna, N., 2014. Assessing uncertainty dimensions in land-use change models: using swap and multiplicative error models for injecting attribute and positional errors in spatial data. International Journal of Remote Sensing, 35 (1), 149–170. DOI: http://dx.doi.org/10.1080/01431161.2013.866293.

Temme, A., and Verburg, P.H., 2011. Mapping and modelling of changes in agricultural intensity in Europe. Agriculture, Ecosystems & Environment, 140 (1), 46–56. DOI: http://dx.doi.org/10.1016/j.agee.2010.11.010.

Thomas, C., et al., 2005. Development of criteria for simplifying ecological risk models. In: MODSIM05 International Conference on Modelling and Simulation, pp. 772-778. From: MODSIM05 International Conference on Modelling and Simulation, 12-15 December 2005, Melbourne, Australia.

Uusitalo, L., 2007. Advantages and challenges of Bayesian networks in environmental modelling. Ecological Modelling, 203 (3–4), 312–318. DOI: http://dx.doi.org/10.1016/j.ecolmodel.2006.11.033.

van der Wel, F. J. M., 2000. Assessment and visualisation of uncertainty in remote sensing land cover classifications. Utrecht: Drukkerij Elinkwijk.

van Horssen, P.W., Pebesma, E.J., and Schot, P.P., 2002. Uncertainties in spatially aggregated predictions from a logistic regression model. Ecological Modelling, 154 (1), 93–101. DOI: http://dx.doi.org/10.1016/S0304-3800(02)00060-1.

van Vliet, J., Bregt, A.K., and Hagen-Zanker, A., 2011. Revisiting Kappa to account for change in the accuracy assessment of land-use change models. Ecological Modelling, 222 (8), 1367–1375. DOI: http://dx.doi.org/10.1016/j.ecolmodel.2011.01.017.

van Vliet, J., et al., 2013. A fuzzy set approach to assess the predictive accuracy of land use simulations. Ecological Modelling, 261, 32–42. DOI: http://dx.doi.org/10.1016/j.ecolmodel.2013.03.019.

Veldkamp, A., and Lambin, E.F., 2001. Predicting land-use change. Agriculture, Ecosystems & Environment, 85 (1), 1–6. DOI: http://dx.doi.org/10.1016/S0167-8809(01)00199-2.

Verburg, P., et al., 2010. Trajectories of land use change in Europe: a model-based exploration of rural futures. Landscape Ecology, 25 (2), 217–232. DOI: http://dx.doi.org/10.1007/s10980-009-9347-7.

Verburg, P.H., and Veldkamp, A., 2004. Projecting land use transitions at forest fringes in the Philippines at two spatial scales. Landscape Ecology, 19 (1), 77–98. DOI: http://dx.doi.org/10.1023/B:LAND.0000018370.57457.58.

Verburg, P.H., et al., 2002. Modeling the Spatial Dynamics of Regional Land Use: The CLUE-S Model. Environmental Management, 30 (3), 391–405. DOI: http://dx.doi.org/10.1007/s00267-002-2630-x.

Verburg, P.H., et al., 2004. Land use change modelling: current practice and research priorities. GeoJournal, 61 (4), 309–324. DOI: http://dx.doi.org/10.1007/s10708-004-4946-y.

Verburg, P.H., Tabeau, A., and Hatna, E., 2013. Assessing spatial uncertainties of land allocation using a scenario approach and sensitivity analysis: a study for land use in Europe. Journal of environmental management, 127, S132-S144. DOI: http://dx.doi.org/10.1016/j.jenvman.2012.08.038.

Verstegen, J.A., et al., 2015. What can and can't we say about indirect land use change in Brazil using an integrated economic - land use change model? GCB Bioenergy. DOI: http://dx.doi.org/10.1111/gcbb.12270.

Walker, R., et al., 2009. Ranching and the new global range: Amazônia in the 21st century. Themed Issue: Land, Labor, Livestock and (Neo)Liberalism: Understanding the Geographies of Pastoralism and Ranching Themed Issue: Organisational Geographies of Power. Geoforum, 40 (5), 732–745. DOI: http://dx.doi.org/10.1016/j.geoforum.2008.10.009.

Walker, W.E., et al., 2003. Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. Integrated Assessment. Integrated Assessment, 4 (1), 5–17. DOI: http://dx.doi.org/10.1076/iaij.4.1.5.16466.

Wallentin, G., and Car, A., 2013. A framework for uncertainty assessment in simulation models. International Journal of Geographical Information Science, 27 (2), 408–422. DOI: http://dx.doi.org/10.1080/13658816.2012.715163.

Warmink, J.J., et al., 2010. Identification and classification of uncertainties in the application of environmental models. Environmental Modelling & Software, 25 (12), 1518–1527. DOI: http://dx.doi.org/10.1016/j.envsoft.2010.04.011.

Wassenaar, T., et al., 2007. Projecting land use changes in the Neotropics: The geography of pasture expansion into forest. Uncertainty and Climate Change Adaptation and Mitigation. Global Environmental Change, 17 (1), 86–104. DOI: http://dx.doi.org/10.1016/j.gloenvcha.2006.03.007.

Wilkinson, L., et al., 2013. An Object-oriented Spatial and Temporal Bayesian Network for Managing Willows in an American Heritage River Catchment. Proceedings of the 2013 Uncertainty in Artificial Intelligence Application Workshops, 77–86.

Wyman, M.S., and Stein, T.V., 2010. Modeling social and land-use/land-cover change data to assess drivers of smallholder deforestation in Belize. Applied Geography, 30 (3), 329–342. DOI: http://dx.doi.org/10.1016/j.apgeog.2009.10.001.

Publikationen

Veröffentlichte Artikel (peer-reviewed)

Krüger, C., and Lakes, T., forthcoming. Revealing uncertainties in land change modeling using probabilities. Transactions in GIS.

Krüger, C., and Lakes, T., 2014. Bayesian belief networks as a versatile method for assessing uncertainty in land-change modeling. International Journal of Geographical Information Science, 29 (1), 111–131. Available at: http://dx.doi.org/10.1080/13658816.2014.949265.

Lakes, T., Müller, D., Krüger, C. (2009). Cropland change in southern Romania: a comparison of logistic regressions and artificial neural networks. Landscape Ecology, 24, 1195-1206

Konferenzbeiträge (Artikel mit peer-review)

Krüger, C., Funke, D., Lakes, T. (2012). Approaching uncertainties in land use change modeling in the Amazon rainforest with Bayesian Belief Networks. In: 6th International Congress on Environmental Modelling and Software (iEMSs). Leipzig, Germany. Inklusive Vortrag

Krüger, C., Funke, D., Lakes, T. (2012). Modeling Spatio-Temporal Patterns of Deforestation in Brazil with a Bayesian Belief Network Approach. In: Geographic Information Zeitgeist (GIZ). Münster, Germany. Inklusive Vortrag

Theisselmann, F., Kühnlenz, F., Krüger, C., Fischer, J., Lakes, T. (2010). Modifying a well-established cellular automata modeling approach to simulate urban land use dynamics: exploring benefits of an experiment management. In: Greve, K., Cremers, A., B. (Eds.) EnviroInfo: Integration of Environmental Information in Europe, 24th International Conference on Informatics for Environmental Protection. Bonn

Lakes, T., John, I., Müller, D., Krüger, C., Rabe, A. (2010). A support vector machine approach to model urban growth in the greater Tirana region, Albania. In: The 13th AGILE International Conference on Geographic Information Science. Guimaraes, Portugal

Konferenzbeiträge (Poster)

Krüger, C. (2013). Uncertainty in spatiotemporal modeling of future land- use scenarios. In: Spatial Statistics Conference 2013. Columbus (Ohio), USA

Eidesstattliche Erklärung

Hiermit erkläre ich, die vorliegende Dissertation selbstständig und ohne Verwendung unerlaubter Hilfe angefertigt zu haben. Die aus fremden Quellen direkt oder indirekt übernommenen Inhalte sind als solche kenntlich gemacht. Die Dissertation wird erstmalig und nur an der Humboldt-Universität zu Berlin eingereicht. Weiterhin erkläre ich, nicht bereits einen Doktortitel im Fach Geographie zu besitzen. Die dem Verfahren zu Grunde liegende Promotionsordnung ist mir bekannt.

Carsten Krüger

Hannover, 11.05.2016